

The Sound Recording Studio

Exploring the Art and Technology of Music Production at Home

Gregory Tait



The Sound Recording Studio © 2023 by [Gregory Tait](#) is licensed under
Creative Commons Attribution-NonCommercial 4.0 International License

Over the past three years, I have written quite a bit on my website about setting up a simple and inexpensive home studio for recording and mixing music. Technical and artistic sound engineering design features are discussed for a computer-based digital recording system, from the sound gear to the digital audio workstation.

As a compilation of these blog posts, this e-book explores the art and technology of music production, from start to finish. It is not a tutorial or 'how to' guide - you can get plenty of that on YouTube and websites. Rather, it strives to provide helpful information on fundamental sound recording topics in a logical sequence. With a better understanding of these fundamentals, I hope readers will gain valuable insights in putting together a music studio that can create some great sound recordings right in their living rooms at home.



About the author

A classically trained pianist, Gregory Tait received his bachelor's degree in music and physics and his doctoral degree in electrical engineering. Drawing on a lifelong involvement in sound engineering, he has set up a home studio for performing and recording musical works. Current projects include recording musical pieces from the classical, Christian, and folk repertoire. He can be contacted at [Pedal Point Sound](#).

Table of Contents

Navigate using the Bookmarks side panel in Adobe Acrobat

Section A: Home Music Studio

- A1. Basic Elements
- A2. The Space
- A3. The Digital Piano
- A4. Audio Cables and Circuits
- A5. Noise
- A6. Unbalanced versus Balanced
- A7. Direct Inject Box
- A8. Audio Signal Levels
- A9. Audio Interface
- A10. Digitization
- A11. Dynamic Range
- A12. Digital Signal Levels
- A13. The Computer
- A14. Latency
- A15. Digital Audio Workstation - Signal Flow
- A16. Monitors
- A17. Sound Source Localization
- A18. Acoustics
- A19. Un-interruptible Power Supply

Section B: Recording

- B1. Recording and Editing
- B2. Digital Audio Workstation - Recording
- B3. Dynamic Microphone
- B4. Ribbon Microphone
- B5. Condenser Microphone
- B6. Microphone Techniques
- B7. Low-Latency Monitoring
- B8. Cue-Mix Monitoring
- B9. Driver Error Compensation
- B10. MIDI Recording
- B11. Audio Repair

Section C: Mixing

- C1. Digital Audio Workstation - Mixing
- C2. Equalization
- C3. Compression
- C4. Reverb
- C5. Fade and Pan
- C6. Mid / Side Signal Processing
- C7. Stereo-ization
- C8. Imaging a Stereo Track

Section D: Mastering

- D1. Mastering Overview
- D2. Mastering - Equalization
- D3. Mastering - Stereo Image
- D4. Mastering - Limiting
- D5. Mastering - Tonal Balance Control
- D6. Export Mixdown

Section E: Recording Project

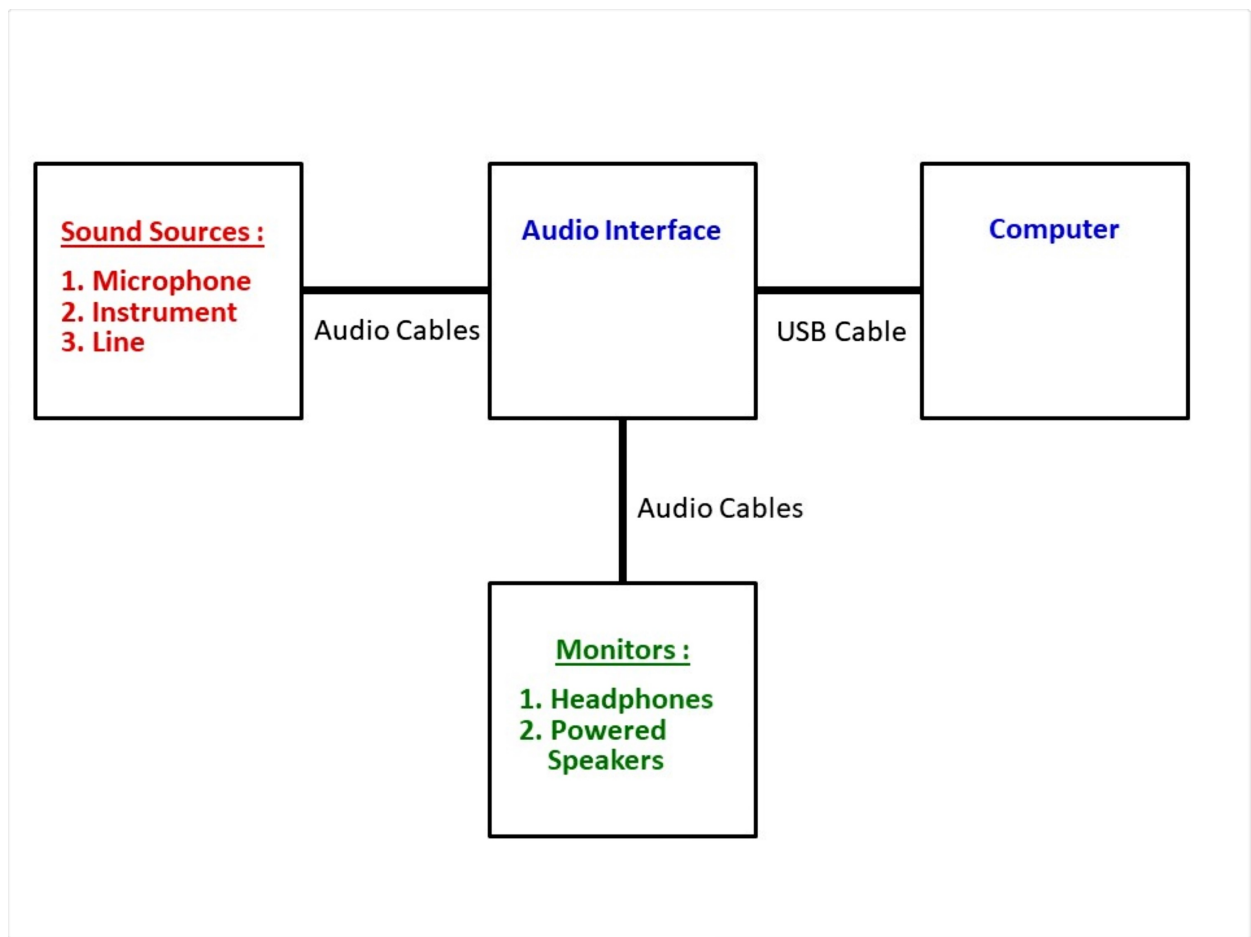
- E1. Session 1: Recording
- E2. Session 2: Mixing
- E3. Session 3: Mastering
- E4. Session 4: Streaming

Appendix: Music Production Education

- AP1. Music Technology Education

A. The Home Music Studio

A1. Basic Elements



The four basic elements of a simple computer-based home recording studio are shown in this block diagram.

1. The analog electrical signals originate from the microphone pickup of sound waves created by a musical instrument and from the direct electrical outputs from instruments such as keyboards and guitars. The "Line" source is the electrical output from playback of a digital tape recorder or mp3 player.
2. The audio interface is the bridge from the analog signal world of the sound gear (microphones, amplifiers, FX electronics, speakers, etc.) to the digital signal world of the computer. Since it is the computer that allows us to set up a home recording studio at a practical cost, the analog-to-digital signal conversion that takes place in the audio interface unit plays a critical role in the system.
3. The computer is the "heart" of the home music studio. Here, the recording, signal processing, mixing, and mastering of the music are accomplished. Computer hardware and software combine to create a digital audio workstation.
4. Playback of the recorded music is done through monitors. Headphones and powered near-field speakers are commonly used in home studios.

In the following chapters, I'll take a closer look at each of these basic elements. Although things may get a bit 'technical' at times, I hope the discussion can shed some light on some of the 'mysteries' of setting up a simple home music recording studio.



A2. The Space



So -- you have to start somewhere. Where are you going to put your studio in your home ?

I chose the 5 ft. x 7 ft. area in a corner of my living room where my old console piano and a small cabinet had been located previously. The space needed to accommodate my new digital piano and piano bench, a work table to hold my sound gear and computer, and a rolling office chair. I planned on using the direct AUX L+R Outputs from the digital piano, so I didn't need to worry about mic'ing (placing microphones) around the piano. There is, however, ample space in the living room next to that 5 ft. x 7 ft. area rug in the photo, should I need to mic a vocalist or acoustic guitar. Additionally, the acoustics in the living room are good - a bright spacious ambient with very little reverberation. The space would not require any sound treatments for recording with microphones. It is also a very quiet and noise-free space -- at least most of the time.

Since this is the *living room* in my family's home, there was a desire that the studio equipment be aesthetically pleasing. Fortunately, the Yamaha CLP-685 piano has an elegant, modern look in matte black finish. Accordingly, a simple modern black brushed-metal work table and black mesh-back chair were needed to complement the piano as part of our living room decor. Here's the "look" we achieved -not bad !



A3. The Digital Piano



For my simple computer-based recording system, there are three essential elements: the sound source (musical instruments), the audio interface, and the computer hardware/software. I'll talk at length about the latter two elements in upcoming posts. Here, my focus is on the "workhorse" instrument - the keyboard. Other instruments, such as guitar, violin, cello, and voice, are also possible sound sources. But since I am a pianist, the keyboard instrument gets full attention here !

I should point out that the choice of specific brands/models for equipment throughout this blog reflects my own preferences, based on functional capabilities and performance of the equipment that I need for my artistic purposes. Furthermore, I am limited in budget, so that has a substantial impact on the equipment selected. I am not endorsing any particular brand/model -- no advertisement is intended !

That said, I was 'sold' on getting a mid-level Yamaha digital piano. In my opinion, Yamaha has the most advanced technology for both sampled sound and mechanical keyboard action. Simply put, I sound better and play better on a Yamaha !

I could afford to acquire a new Yamaha Clavinova CLP-685. This piano is the least expensive model that has all the advanced technology I wanted in its grand piano voicing and action, including binaural sampling of the Yamaha CFX grand piano, the touch and release (escapement) of the counter-weighted keys and the 88-key linear graded hammers. That's my Yamaha Clavinova in the photo above.

In addition to being my grand piano, the CLP-685 provides hundreds of other synthesized voices that can serve as my 'virtual instruments' for sound sources. While the keyboard comes with a MIDI (musical instrument digital interface) output that can be used to record instrument tracks on the computer, I solely use the analog audio signal output of the CLP-685. I prefer to capture the full dynamic nature of a 'live' performance in the analog signal that is recorded to audio tracks on the computer.

The analog signal from the Yamaha keyboard instrument is output via two 1/4" TS

(Tip-Sleeve) jacks - a left channel and a right channel. These analog signal connections, shown as "AUX OUT" in the photo below, require the appropriate instrument cables to connect to the next piece of audio gear in the recording chain. I will talk a lot about cables and connectors in the next session.



A4. Audio Cables and Circuits

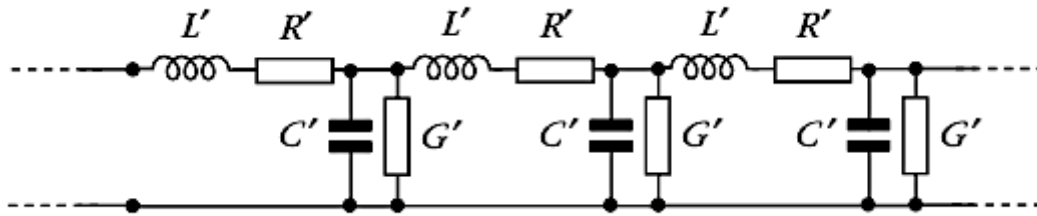


Now we're ready to connect our sound source (microphone, pickup, instrument output) to the next 'building block' in our chain of signal processing, usually a mixer console or audio interface unit. To do this, we 'transmit' our audio signal through a cable connected between the sender and receiver. This sounds like a very mundane task. But there is definitely an art and science to this, which warrants taking a closer look at cables and connectors. There are many blogs and videos on this topic -- not all saying the same thing ! For the next few posts, I'll put on my electrical engineering hat to add my two cents to the mix using electronic circuit theory.

Cable Equivalent Circuit

Cables used for transmitting signals consist of long lengths of metal wires in either a two-conductor or a three-conductor wire configuration. These wires are part of the full electrical circuit that starts with the

source electronics and ends with the receiver electronics. The cable is modeled by the circuit elements in the figure below ,



J.C.G. Lesurf St. Andrews University

The circuit elements are not lumped components but are distributed quantities 'spread' along the long lengths of the metal conductor system. The inductance per meter L' models the effect of the time-changing current in the wire (changing magnetic flux), and the capacitance per meter C' models the effect of the time-changing potential difference (voltage) between the metal conductors (changing electric field). The resistance per meter R' models the small resistive loss (Joule heating) due to the metal's finite conductivity, and the conductance per meter G' models the small polarization loss of the dielectric material that separates the conductors.

When the signal voltage varies at very high frequencies (radio frequencies $f > 100$ MHz), the cable acts as a transmission line, guiding the voltage waves down its length. The wavelength λ of the voltage wave is given by

$$\lambda = v/f$$

where v and f are, respectively, the speed and frequency of the wave. The speed of an electromagnetic wave in a coaxial cable filled with dielectric material is about 2/3 the speed of light, or $\sim 2 \times 10^8$ m/s.

For frequency $f = 1$ GHz, the wavelength $\lambda = 0.2$ m ≈ 8 inches. So the wave nature of the transmission line would have to be accounted for in a circuit analysis when interconnects are longer than a couple of inches. Voltage and current waves travel forward and backward on this transmission line, reflecting at its terminations where there is an impedance mismatch. There exists a characteristic impedance (Z_c) for this high-frequency transmission line given by the ratio of the amplitude of the forward-propagating voltage wave (V^+) to the amplitude of the forward-propagating current wave (I^+),

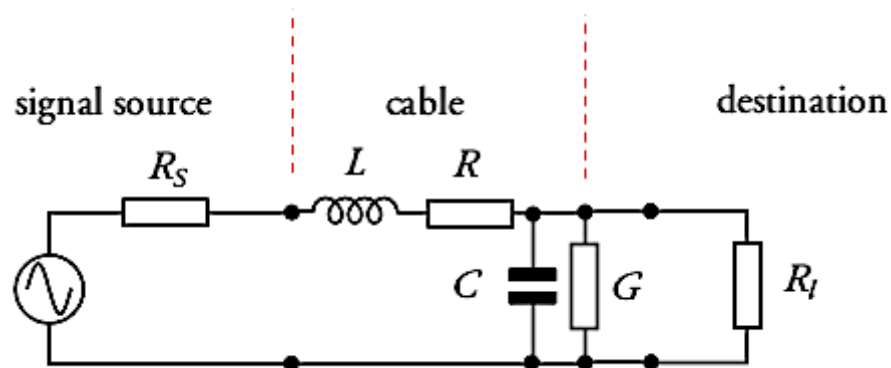
$$Z_c = \frac{V^+}{I^+} = \sqrt{\frac{L'}{C'}}$$

where L' and C' are, respectively, the distributed inductance and capacitance of the transmission line. L' and C' are determined by the configuration and geometry of the cable and by the dielectric material properties of the cable. Cables are designed to have standardized characteristic impedances -- 75 Ohm (coax) and 300 Ohm (twin-lead) cables for UHF/VHF applications, and 50 Ohm (coax) cables for radio-frequency and microwave circuit applications. Please note that if you look at a cable, you won't see a 75 Ohm, 300 Ohm or

50 Ohm resistor anywhere !! The characteristic impedance Z_c has units of Ohms since it is defined as the ratio of voltage amplitude to current amplitude in the wave propagating down the transmission line. It is a useful quantity in the analysis of high-frequency circuits.

At audio frequencies, the wavelength $\lambda > 10,000$ m, which is about 6.2 miles !!

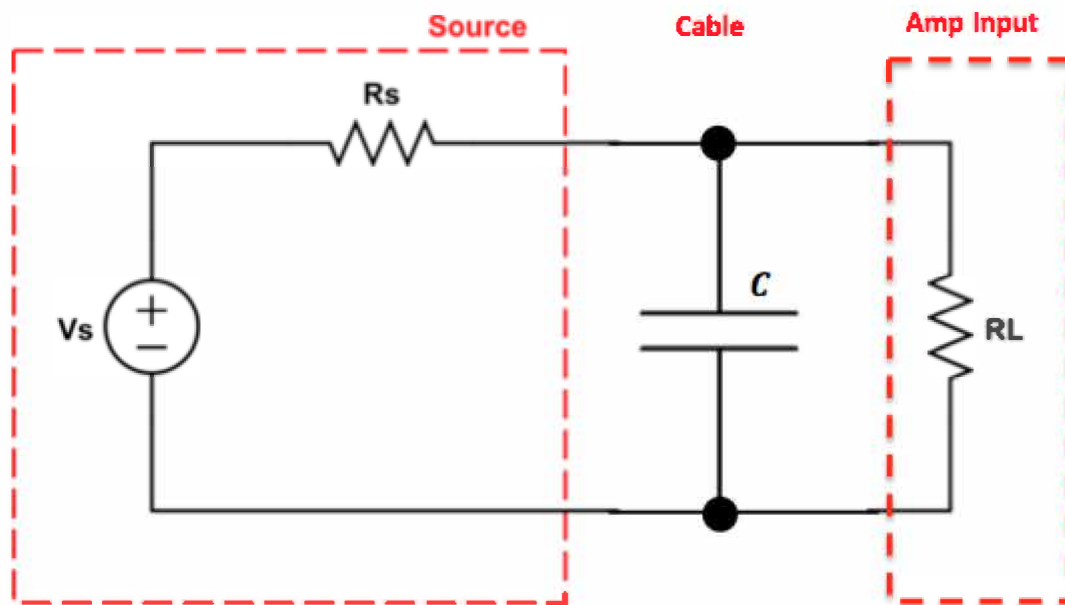
SO in the equivalent circuit model for an audio frequency cable, we really can consider the distributed circuit elements of the transmission line to be simply lumped components.



J.C.G. Lesurf St. Andrews University

Furthermore, the inductive reactance ($2\pi fL$), metal resistive loss (R) and dielectric loss (G) are quite small in audio cables, so these elements will be neglected here. The primary feature of the audio cable is its capacitive loading effect on the circuit. This loading effect increases with frequency and with cable length. In essence, the audio cable is nothing more than wires connecting together two circuits – just like the wires you use on a breadboard to build electronic circuits. But the wires of the audio cable do present a sizeable parasitic capacitance in the circuit.

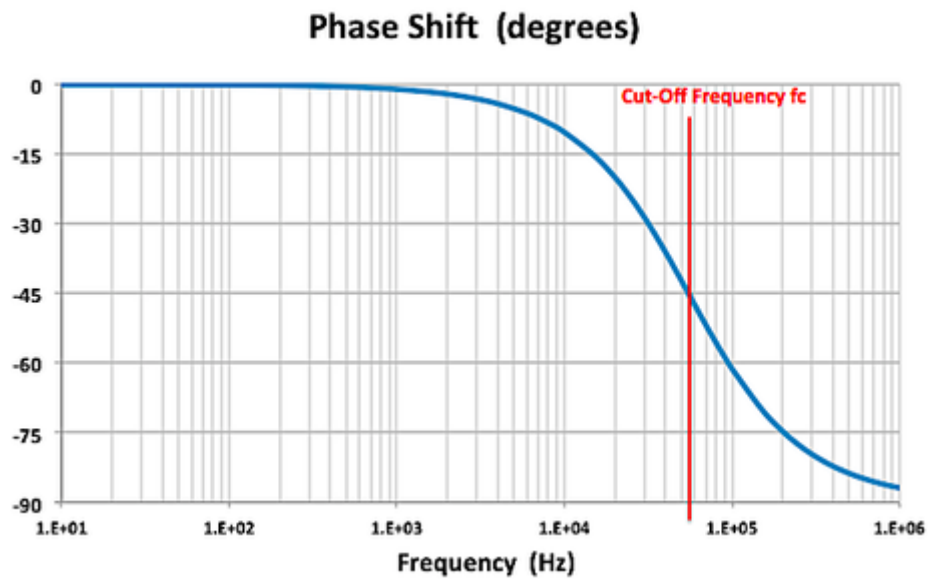
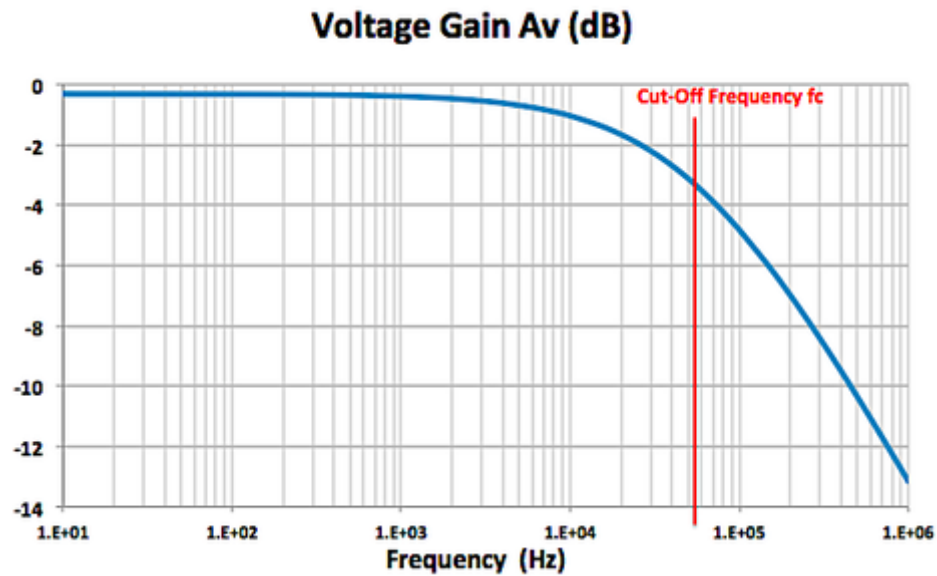
The capacitive loading on the circuit creates a low-pass RC filter. The source resistor R_S is the Thevenin equivalent output impedance of the signal source circuit. The load resistor R_L is the input impedance looking into the input of the next stage (usually an amplifier circuit). The amplitude and phase changes of the output sinusoidal voltages $V_{out} = V_{RL}$ from the input sinusoidal voltages $V_{in} = V_S$ are calculated using AC circuit analysis.



Voltage Gain: $A_v = \frac{|V_{out}|}{|V_{in}|} = \frac{R_L / (R_L + R_S)}{\sqrt{1 + (2\pi f C (R_S || R_L))}}$

Phase Shift: $\theta = \tan^{-1} [-2\pi f C (R_S || R_L)]$

Cut-off Frequency: $f_c = \frac{1}{2\pi (R_S || R_L) C}$, $R_S || R_L = \frac{R_S \cdot R_L}{R_S + R_L}$



At frequencies well below the cut-off frequency f_c , the amplitude V_{out} is given by the voltage divider of expression,

$$|V_{out}| = R_L / (R_L + R_S) |V_{in}|$$

Also, for frequencies well below f_c , there is only very small phase shift of the sinusoidal voltage from input to output ,

$$\theta \cong 0 \text{ degrees}$$

The audio signal is a voltage signal that is processed through electronic circuits from the sound source to the analog-to-digital (A/D) converter, and from the digital-to-analog (D/A) converter to the power amplifier driving the monitor speakers. As the signal is passed from one electronic circuit to the next, we need to pay attention to this voltage divider action. The Thevenin equivalent circuit of the sound source is comprised of the voltage source V_S and the output resistance $R_{out} = R_S$. The input of the next circuit, such as the pre-amp in the mixer board or audio interface box, is comprised of its input resistance $R_{in} = R_L$, i.e., the input resistance of the second circuit is the load resistance of the first circuit. Therefore, to maintain signal strength (voltage amplitude), we can see from the voltage divider equation above that we desire to have a sound source with a low output resistance (low impedance or Lo-Z) and a pre-amp with a high input resistance (high impedance or Hi-Z). And this is exactly what is done in sound gear. Impedance levels are roughly categorized as follows:

Low Impedance:	less than 600 Ohms
Medium Impedance:	600 Ohms – 10k Ohms
High Impedance:	greater than 10k Ohms

Complex time-domain voltage signals will contain high-frequency components. As the frequencies of these components approach the cut-off frequency of the audio cable's low-pass filter, say above $\sim f_c/10$, the amplitude and phase shift of each component will be altered. This means that the time-domain voltage waveform at the load will not be an exact duplicate of the waveform at the source -- this is DISTORTION of the signal!

As an example, a 100-ft. long audio cable may have a capacitance on the order of $C=2$ nF. The source (microphone) has a low output resistance $R_S = 150$ Ohm, and the mixer pre-amp has a relatively high input resistance $R_L = 4k$ Ohm. This gives a cut-off frequency

$$f_c = \frac{1}{2\pi (R_S || R_L) C} \cong 550 \text{ kHz}$$

A sound waveform with fundamental frequency of 10 kHz may have harmonics at and above 50 kHz that will be affected by the low-pass filtering of the cable. It may be, however, that the distortion created in the time-domain signal waveform in this example is inaudible to humans. If an audio cable with capacitance $C = 20$ nF had been used in the above example, the cut-off frequency would be $f_c \cong 55$ kHz and distortion of the sound would be significant and quite audible. We should always be mindful of keeping source resistance and cable capacitance to low values.

In the next chapters, I'll take a look at noise in our audio circuits and ways to control it, including the use of balanced audio cables, impedance transformers (DI boxes) and ground shields.

A5. Noise in Audio Circuits



Before continuing the discussion of hooking up a sound source to the pre-amplifier of a mixer console or audio interface, I want to bring up the topic of noise – the unwanted sounds in our recordings.

The description, analysis and mitigation of noise in electronic circuits are vast subjects covered in textbooks, research papers, and university engineering courses. I'll just touch on the more pertinent information here. For an expert treatment on this subject, Henry Ott's seminal monograph is probably the best source of information:

Henry Ott, "Noise Reduction Techniques in Electronic Systems," 2nd Ed., Wiley, 1988.

There are many sources of noise that can inject a noise voltage into our circuit and 'interfere' with our desired signal voltage, giving us a degraded sound recording. These sources of noise can be categorized as being internal to the circuit and external to the circuit.

Internal Sources:

1. Thermal Noise

This noise originates from the random thermal motion of electrons in conducting materials, and produces a root-mean-square (rms) voltage

$$V_n = \sqrt{4 K T B R_s}$$

where,

$K = 1.38 \times 10^{-23} \text{ J/}^\circ\text{K}$ Boltzmann's Constant

$T = 300 \text{ K}$ room temperature

$B = 20 \text{ kHz}$ (audio bandwidth)

$R_s =$ source resistance (Ohms)

This noise is primarily associated with resistors. For a source resistor $R_s = 150 \text{ Ohms}$, the thermal noise voltage is approximately $V_n = 223 \text{ nV}$.

2. Shot Noise

This noise originates from the discrete nature of electric charge and its emission across potential barriers in semiconductor devices like diodes and bipolar junction transistors. Although it is usually small compared to thermal noise, shot noise is a contributor to the noise added to a signal by amplifier circuits that contain many diodes and bipolar junction transistors.

3. Flicker or "1/f" Noise

This noise is a waveform whose power spectral density (PSD) goes down with increasing frequency f with a $1/f$ dependence -- this is characterized as "pink noise". The origin of this noise in electronics is most likely due to the impurity traps and generation/recombination centers in semiconductor materials used for the diodes and transistors.

External Sources:

1. Ambient Noise

This is the background noise picked up by a microphone, if a microphone is being used.

2. Electromagnetic Interference

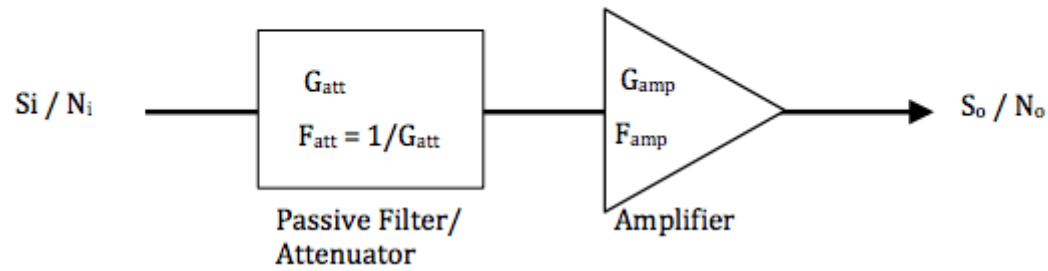
The external electric and magnetic fields present in the space surrounding our circuit can be coupled to the metal wires in our circuit. In particular, the long wires of our audio cable can act as an antenna in receiving electromagnetic radiation. The result is 'noise' (unwanted) currents and voltages induced in our circuit. Sources of this electromagnetic interference range from the low-frequency (60 Hz) magnetic fields from AC power lines to the high-frequency electric fields of radio and microwave radiation.

All of the noise sources described above create uncorrelated noise. When noise voltages are produced independently and there is no relationship between their instantaneous amplitudes or phases, they are said to be uncorrelated. Total noise power is then the sum of individual noise powers. Therefore, the resultant total noise voltage is the square root of the sum of the squares of the individual noise voltages.

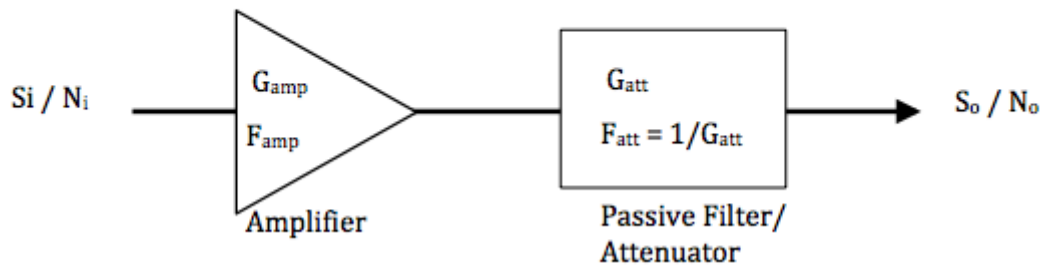
There is also noise categorized as correlated noise. Correlated noise voltages are related to the signal voltages, and are caused by nonlinear current-voltage relationships in active devices such as the transistors in amplifiers. Nonlinear amplification of signal voltages leads to the production of unwanted harmonic voltages (mixing of signals at the same frequency) and intermodulation voltages (mixing of signals at different frequencies). These two nonlinear effects are quantified by the amplifier specifications of Total Harmonic Distortion (%THD) and Intermodulation Distortion (usually specified by a quantity called the third-order intercept point IP3).

The presence of noise from internal sources in our circuit is unavoidable – it's just the physics of the motion of electrons in conducting and semi-conducting materials. But there are things we can do to try to control and mitigate noise through good circuit design practices. A good example of this is to get a low-noise amplifier into the signal path as soon as possible, especially if the sound source has a low signal-to-noise power ratio (S/N), such as with a microphone. To show this, consider these two alternative circuits:

1. Filter/Attenuator – Amplifier Circuit:



2. Amplifier – Filter/Attenuator Circuit:



where,

G and F are, respectively, the power gain (unitless) and noise factor (unitless) of the circuit blocks. The noise factor F_{amp} is given by,

$$F_{amp} = \frac{N_i + N_{amp}/G_{amp}}{N_i} > 1$$

and serves as a figure of merit for how much noise power N_{amp} is added to the amplifier output by the electronics inside the amplifier. In the spec sheet for an audio amplifier, its noise performance is given by the Equivalent Input Noise Voltage EIN (V), which is related to noise factor F as,

$$EIN = \sqrt{4KTBR_s} \cdot \sqrt{F} \quad [V]$$

A low-noise amplifier has a noise factor $F \sim 2$, and with a source resistor $R_s = 150$ Ohms, the corresponding equivalent input noise voltage per unit bandwidth ($B = 1$ Hz) would be

$$EIN = \sqrt{4KTR_s} \cdot \sqrt{F} = 2.23 \text{ nV}/\sqrt{\text{Hz}}$$

For the circuits above, the overall noise factor for the cascade of the two circuit blocks together can be estimated from the Friis formula ,

1. Filter/Attenuator – Amplifier Circuit:

$$F_{F-A} = \frac{S_i/N_i}{S_o/N_o} = F_{att} + \frac{F_{amp} - 1}{G_{att}} = \frac{F_{amp}}{G_{att}}$$

2. Amplifier – Filter/Attenuator Circuit:

$$F_{A-F} = \frac{S_i/N_i}{S_o/N_o} = F_{amp} + \frac{F_{att} - 1}{G_{amp}} = F_{amp} + \frac{1/G_{att} - 1}{G_{amp}}$$

where, $G_{amp} \gg 1$ and $G_{att} < 1$

The noise factor of an amplifier circuit depends on the types of active semiconductor devices used, as well as on the embedding impedances and operating conditions (bias currents) of the active devices. Although more costly, it is worthwhile to use a very-low-noise amplifier with high gain up front in a cascade of circuits, since the overall noise factor of the combined circuits will be dominated by the noise factor of this front-end amplifier. As an example, suppose we had the following:

Amplifier: $F_{amp} = 1.4$ $G_{amp} = 100$ (+20 dB power gain)
 Filter/Attenuator: $G_{att} = 0.5$ (-3.0 dB power loss)

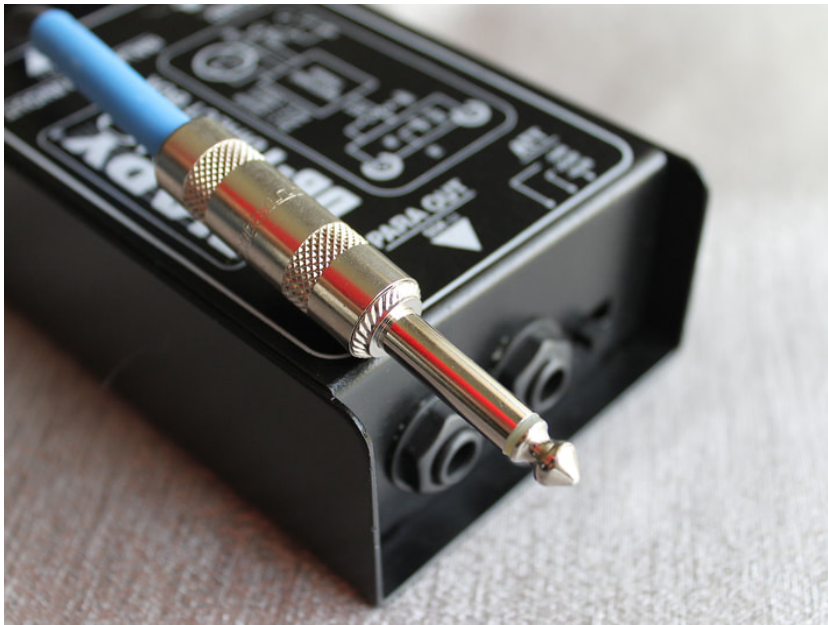
$$F_{F-A} = \frac{F_{amp}}{G_{att}} = \frac{1.4}{0.5} = 2.8 \quad (4.5 \text{ dB})$$

$$F_{A-F} = F_{amp} + \frac{1/G_{att} - 1}{G_{amp}} = 1.4 + \frac{2 - 1}{100} = 1.41 \quad (1.5 \text{ dB})$$

Clearly, having the amplifier precede the filter/attenuator yields a 3 dB improvement in output S/N ratio. Keeping noise under control in each mixer channel is important because the noise power of each channel will sum together in the output bus.

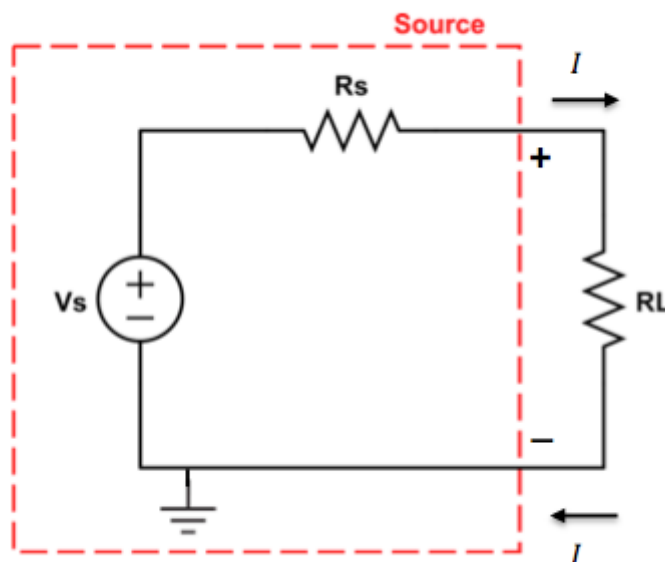
Mitigating the effect of external electromagnetic interference on the circuit will be the subject of my next chapter as I look at the difference between unbalanced and balanced audio cables.

A6. Unbalanced vs. Balanced



In this chapter, I return to the subject of **audio cables** and hooking up the output of a sound source to the pre-amplifier of a mixer console or audio interface.

The electrical circuit consisting of the sound source, the audio cable and the load (the input of the pre-amp) is shown in its most simple form in the following figure.



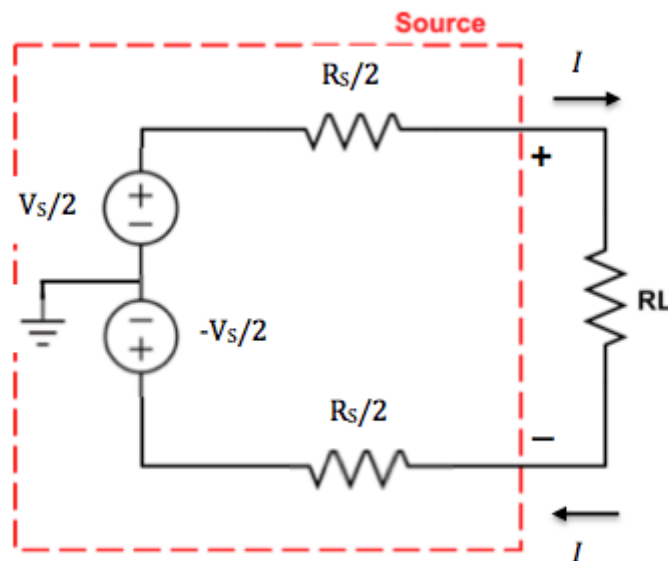
The Thevenin equivalent circuit for the sound source consists of a time-varying voltage $V_s(t)$ and an output resistance R_s (or more generally, an output impedance $Z_s = R_s + jX_s$, where the reactance X_s originates from capacitance and inductance in the source). The audio cable is simply the two metal conductors connecting the positive (+) terminal (or “red wire”) and the negative (-) terminal (or “black wire”) of the source to the load. A high conductivity metal wire at low frequencies is a surface of constant potential, so the potential difference (voltage) between the two wires remains the same all the way down the cable from source to load. Of course, this voltage along the cable is varying with time – it is the signal voltage $V_{\text{SIGNAL}}(t)$. The load resistance R_L is the equivalent input resistance of the pre-amp (again, more generally, an input impedance $Z_L = R_L + jX_L$). In the circuit above, the negative lead (black wire) is shown connected to an earth ground. This is done to give a reference potential (0 V) to a point in the circuit, so that the potentials in the circuit are not “floating”. Grounding the circuit is a good safety practice for AC-powered electronic circuits, and as we will see, is also useful in providing some noise immunity. The point at which the circuit is grounded is entirely arbitrary, but is commonly done on the negative lead wire at the source.

It should be noted that we are focused on voltages in an audio circuit – our information signal is a voltage that is ‘processed’ through a chain of circuits, like voltage amplifiers, filters and transformers. Voltage doesn’t “flow” in a circuit – it is simply the potential difference between two points in the circuit. We don’t usually focus on the low-level current $I(t)$ that does flow in the circuit, but it is there. In fact, the word “circuit” means a closed loop path, without which we would not have a functioning electrical system. The same current I shown leaving the positive terminal of the source in the figure above *returns* to the source via the negative terminal. Even though the negative wire may be grounded, this wire is a necessary part

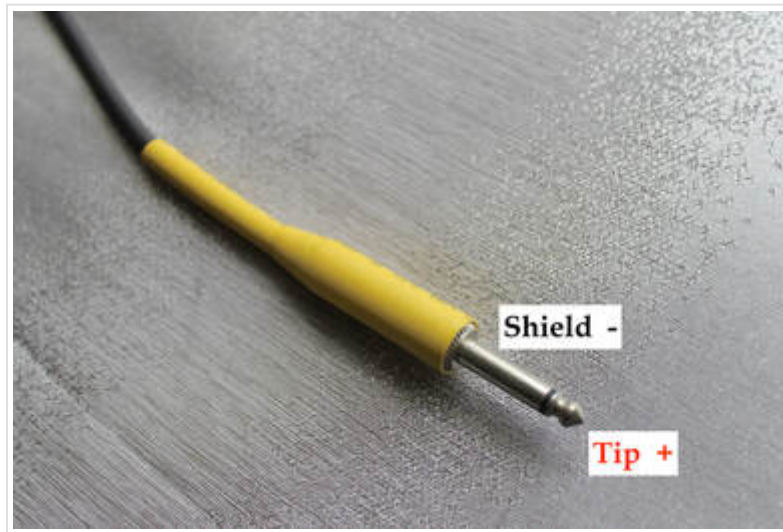
of the signal path.

[In a future post, I'll talk about the circuit formed by the power amp – speaker cable – speaker (monitor). Here, *both* voltage and current need to be considered, as what is desired is to get the maximum **power** transferred from amp to speaker.]

It was mentioned above that the placement of the ground in the circuit is arbitrary. In the circuit below, a different choice is made for the location of the ground. But the operation of this circuit is absolutely identical to that of the original circuit above. The load resistor can't "tell" the difference! But these two circuits are key to understanding the difference between an unbalanced audio cable and a balanced one.



A two-wire cable connected to the positive (+) and negative (-) terminals of the source in the first circuit configuration above is said to be "unbalanced", since the signal voltage is maintained on the positive wire with respect to the grounded negative wire. The connectors used on the ends of this two-wire unbalanced audio cable are typically 1/4-inch TS (Tip-Shield) plugs. The Tip is connected to the insulated positive wire that runs inside the cable. The Shield is connected to the grounded negative wire that is a metal braided mesh or foil surrounding the insulated inner wire.



2-wire unbalanced cable with 1/4" TS connector

A three-wire cable connected to the positive (+), negative (-), and ground terminals of the source in the second circuit configuration above is said to be “balanced”, since the signal voltage is maintained between the positive and negative wires and is symmetrically centered around ground potential (0V). The connectors used on the ends of this three-wire balanced audio cable are typically XLR 3-pin male plugs and female jacks. Pins 2 and 3 in the XLR shell are connected to the insulated positive and negative wires, respectively. Pin 1 of the XLR is connected to the outer metal shield conductor surrounding the insulated inner wires. A balanced cable can also be terminated with a 1/4-inch TRS (Tip-Ring-Shield) connector. (It should be pointed out that a 3-wire cable with 1/4-inch TRS connectors is also frequently used as a stereo two-channel (left and right) unbalanced cable, such as your headphone cable.)



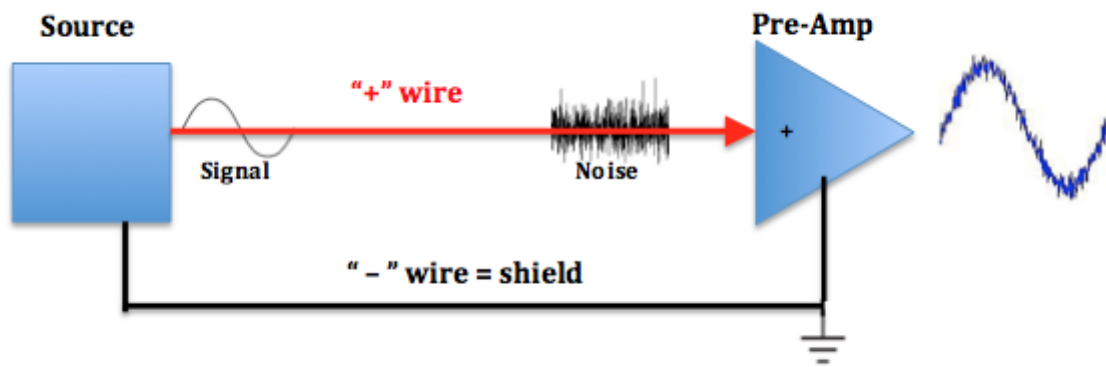
3-wire balanced cable with XLR connector



3-wire balanced cable with 1/4" TRS connector

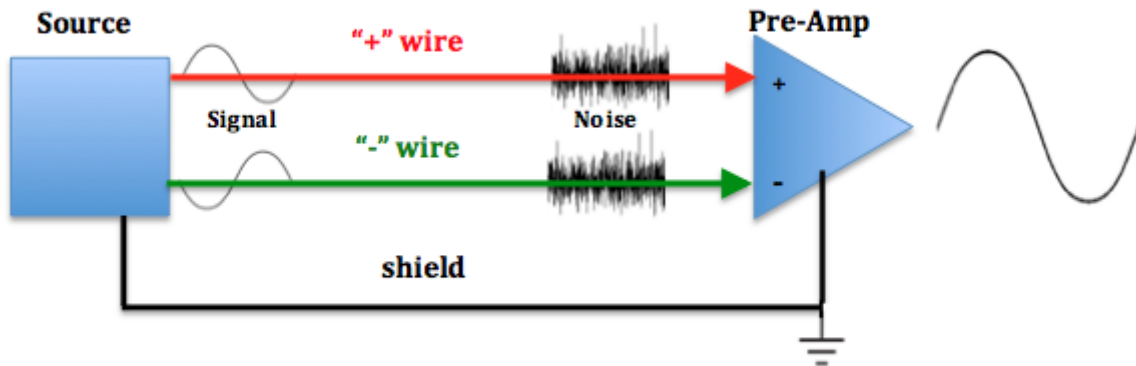
At this point, please note that unbalanced and balanced audio cables should NOT be called high impedance (Hi-Z) and low impedance (Lo-Z) cables, respectively. These cables are NOT high-frequency **wave-guiding transmission lines** that have a characteristic impedance Z_c . These cables are just wires that connect sources to loads, and it is the source and load that are described by high and low impedances. It has become common practice to associate XLR balanced cables with microphones that are low-impedance sources – hence the erroneous attribution of low impedance to the XLR balanced cable. Similarly, it is common practice to associate 1/4" TS unbalanced cables with guitar pick-ups that are high-impedance sources – hence the erroneous attribution of high impedance to the 1/4" TS unbalanced cable.

The use of balanced audio cables becomes necessary when you have a low-voltage (low S/N) source, such as a microphone, that needs to be connected to a pre-amp that is a long distance away. In such cases, the mitigation of noise induced on the wires from external electromagnetic interference is critical. The outer grounded metal conductor is partially effective in shielding the inner wires from electromagnetic interference. However, currents induced in the outer conductor from electric fields can penetrate through the braid and around the connectors to the interior of the cable, and low-frequency magnetic fields penetrate the shield easily. In an unbalanced cable, the noise voltage created on the positive wire is added to the signal voltage, and both are amplified in the single-ended pre-amp, as shown below.

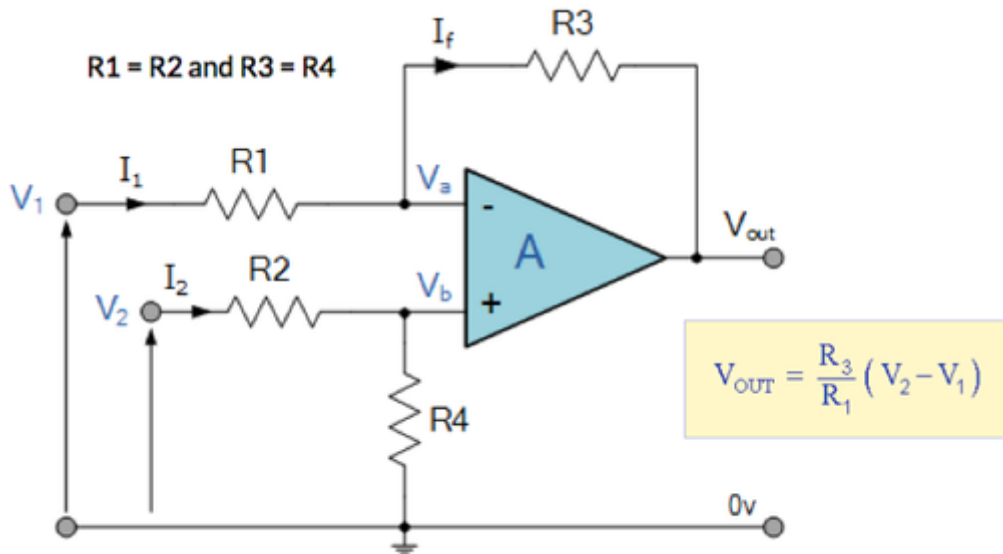


Unbalanced Circuit

In a balanced cable, noise voltage (with respect to ground potential) is induced equally on the co-located positive and negative wires. By using a differential input amplifier, however, the common-mode noise voltages can be suppressed, as shown below.



Balanced Circuit



$$+ \text{ wire : } \quad V_2 = +\frac{V_{signal}}{2} + V_{noise}$$

$$- \text{ wire : } \quad V_1 = -\frac{V_{signal}}{2} + V_{noise}$$

$$\begin{aligned} V_{out} &= A_V(V_2 - V_1) = A_V \left[\left(+\frac{V_{signal}}{2} + V_{noise} \right) - \left(-\frac{V_{signal}}{2} + V_{noise} \right) \right] \\ &= A_V \cdot V_{signal} \end{aligned}$$

where A_V is the amplifier voltage gain.

In practical differential-input amplifier circuits, the transistor devices are not perfectly matched and there are imbalances in bias currents. So there isn't perfect cancellation of the common-mode noise voltages. Nonetheless, there is significant suppression of noise, and the use of balanced audio cables with differential-input amplifiers is well worth the effort.

Next time, we'll take a look at how to transform an unbalanced circuit into a balanced one – the role of the Direct Inject (DI) box.

A7. Direct Inject Box



OK, the technical talk continues – another post in our series of connecting a sound source to the input of a pre-amplifier in a mixer console or audio interface.

The goal all along has been to get a good portion of the signal voltage generated in the source transferred to the input of the pre-amplifier, and to do so without introducing too much noise into the system and without distorting the signal waveform.

We've seen that it is desirable to have **low source impedance and high load impedance** due to the voltage divider effect of the circuit. Along with low source impedance, it is desirable to have a cable with as little parasitic capacitance as possible so that there is little high-frequency distortion of the signal waveform due to the low-pass filtering effect of the cable.

So, cable lengths should be as short as practical. Additionally, a balanced cable with a grounded shield can have less capacitance per unit length between the positive and negative wires than an unbalanced cable. Lastly, to keep noise to a low level, particularly in long cable runs from source to load, we need to use a **three-wire balanced audio cable** with a properly grounded shield and a **low-noise (low EIN voltage)** differential-input pre-amplifier.

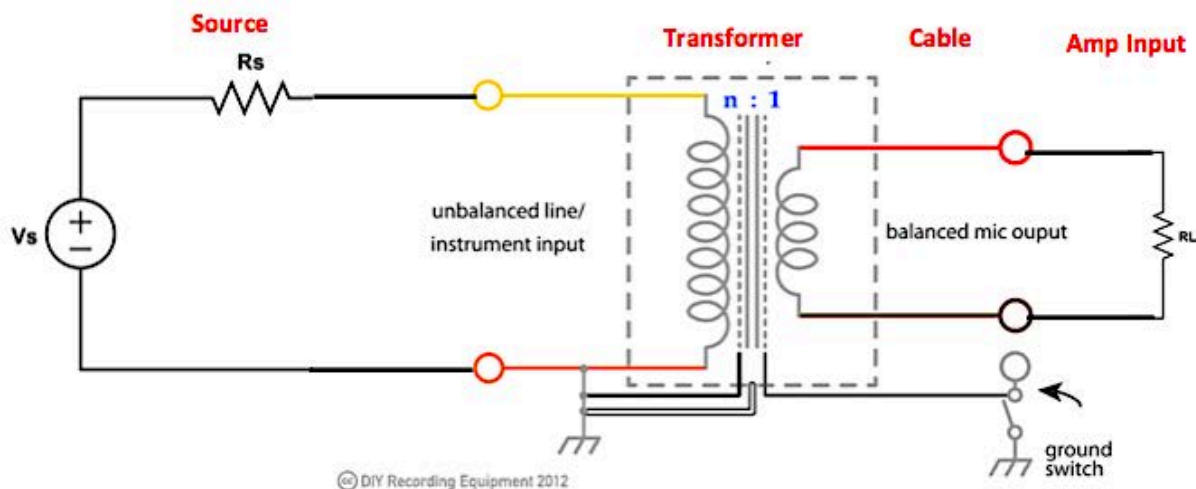
The microphone input impedance of a modern pre-amp is typically on the order of 2k – 10k Ohms. The source impedance of a good microphone is on the order of 150 Ohms (with some vintage, and still very popular, models such as the Shure SM57 being about 600 Ohms). These microphones typically have an XLR connector, making them suitable to attach directly to a balanced audio cable that goes to the XLR connector on the mixer console or audio interface.

In contrast, electric guitars and piezo pickups on acoustic guitars have source impedances that are relatively high, in the range 5k – 20k Ohms. Furthermore, output connections from these instruments are typically ¼" TS jacks, meant for attaching unbalanced instrument cables. This is a “double whammy” – high source impedance and an unbalanced instrument cable. Often, there is a Line/Instrument input (¼" TS jack) on the mixer or audio interface that has a built-in pad (voltage attenuator) that boosts the input impedance up to 10k – 20k Ohms. This higher input impedance presents a high enough impedance to the guitar instrument so that the pickup circuit is properly loaded, i.e., does not have to source more current than it is capable of providing. So connecting directly to this Line/Instrument input is acceptable, as long as the instrument cable length remains short ! But in practice, there is usually a long distance between instrument and the sound board.

What to do ? The magic box – the Direct Inject (DI) box – comes to the rescue ! The DI box can accomplish the following tasks:

1. Present a high-impedance load (> 100k Ohms) to the instrument source, so that it is properly loaded.
2. Provide a low-impedance, balanced output suitable for attaching a balanced cable to run all the way to the mixer console microphone XLR input connection.
3. Electrically isolate the source circuit from the load circuit. The two circuits are magnetically coupled, and each can have its own ground.

The primary electrical component in a DI box is a transformer – two wire coils that share **time-varying** magnetic field flux (no direct-current connection).



The ratio of the number of loops (turns) of the primary coil (the one on the left) and the secondary coil (the one on the right) is $n : 1$, where $n > 1$ is an integer.

The impedance R_1 looking into the primary coil is what the source “sees” as the load impedance,

$$R_1 = n^2 R_L$$

and the voltage V_1 across the primary coil, which is referenced to the local chassis ground, is given by,

$$V_1 = \frac{R_1}{R_1 + R_S} V_S$$

The current I_1 flowing into the primary coil is very small, since R_1 is large.

$$I_1 = \frac{V_S}{R_1 + R_S}$$

The impedance R_2 looking into the secondary coil is what the load “sees” as the source impedance,

$$R_2 = \frac{R_S}{n^2}$$

which is now effectively a low-impedance source, as we desire.

The voltage V_2 across the secondary coil (which is the same as the voltage V_L across the load) is given by,

$$V_2 = \frac{V_1}{n} = \frac{R_1/n}{R_1 + R_S} V_S = \frac{n R_L}{n^2 R_L + R_S} V_S = V_L$$

This signal voltage $V_2 = V_L$ is maintained between the positive and negative wires inside the balanced cable and is symmetrically centered around ground potential (0V). The outer shield of the cable is grounded at the far end of the cable at the pre-amplifier circuit.

Let's put some example values into these equations to see how the DI box transforms impedances.

Source impedance (guitar pickup):
Load impedance (pre-amp input impedance):
Transformer turns ratio $n:1$:

$R_S = 10\text{k Ohms}$
 $R_L = 2\text{k Ohms}$
 $n = 8$

$$R_1 = 128\text{k Ohms}$$
$$V_1 = 0.928 V_S$$

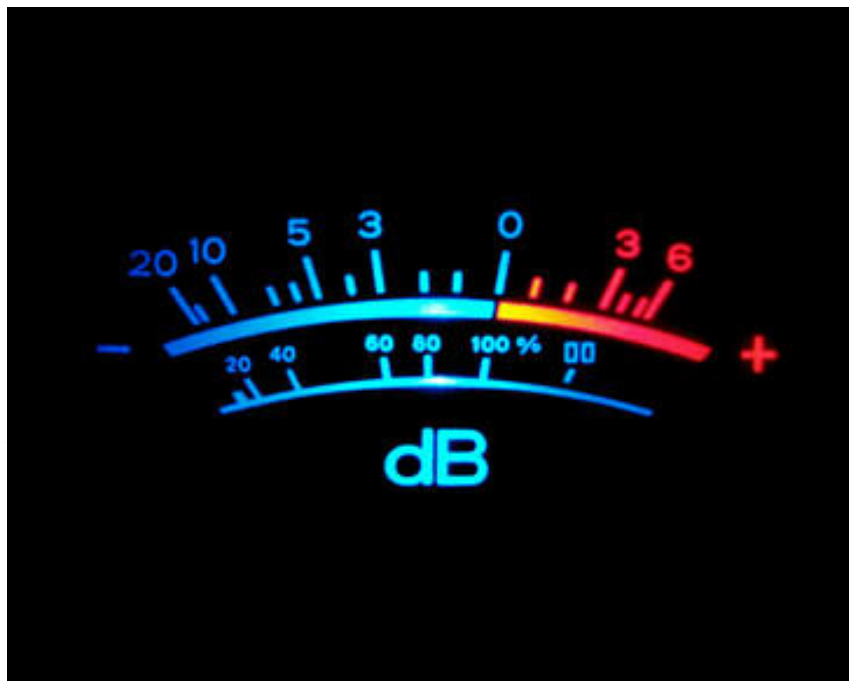
$$R_2 = 156\text{ Ohms}$$
$$V_2 = 0.116 V_S$$

The transformer presents the desired high-impedance (128k Ohms) load to the unbalanced instrument source, while also providing an effective low-impedance (156 Ohms) source connected to the pre-amp microphone input via a fully balanced cable.

A note on signal voltage levels is warranted here. The voltage delivered to the pre-amp input with the transformer in the circuit is, from above, $V_L = 0.116 V_S$, which is significantly reduced from the value V_S (-18.7 dB reduction). This occurs because the transformer is configured to transform impedances to the proper levels, and with its primary-to-secondary turns ratio of 8:1, the voltage is actually “stepped down”. In most cases this is not a problem, because instrument output levels range from tens of millivolts all the way up to a couple of volts (line level output from electronic keyboards). So the reduced voltage level is now commensurate with microphone voltage levels and, hence, is suitable for connecting to the microphone input of the pre-amplifier at the mixer board.

A last word regarding the DI box – about the “ground lift” switch. The source circuit and cable/load circuit are electrically isolated from each other. They are coupled via the time-varying magnetic field flux passing through both primary and secondary windings of the transformer. The source circuit is shown tied to its own local chassis ground (which may be connected to another ground through the shield of the unbalanced instrument cable). The balanced cable shield is grounded at the pre-amplifier circuit. If the chassis ground is isolated from earth ground at the DI box, then connecting the DI box chassis to the earth ground of the balanced cable shield should be OK, as there is a single-point connection to earth ground. However, if the DI box chassis is connected to its local earth ground (e.g., via a connection to the output jack of an electronic keyboard), then there would exist a two-point connection to earth grounds that have slightly different electric potentials. This forms a ground loop circuit carrying noise currents that can create a noticeable ‘hum’ or ‘buzz’ in the sound. To avoid this situation, the ground lift switch is opened and the ground loop circuit is broken.

Audio Signal Levels



At this point, I need to remark on the signal voltage levels that exist through the analog circuit , from source to the analog-to-digital converter and from the digital-to-analog converter to the monitor speakers. The units of measurement for signal voltages were standardized back in the early days of telephone – the deciBel (dB) – so named in honor of Alexander Graham Bell. This is a logarithmic unit based on the fact that the human ear hears sound pressure levels on a logarithmic scale (more about this when we talk about speakers). The following definitions apply:

$$Voltage (dBV) = 20 \log \left(\frac{Voltage (V)}{1 V} \right)$$

$$Voltage (dBu) = 20 \log \left(\frac{Voltage (V)}{0.775 V} \right)$$

$$Power (dBm) = 10 \log \left(\frac{Power (mW)}{1 mW} \right)$$

$$Gain (dB) = 20 \log \left(\frac{V_{out} (V)}{V_{in} (V)} \right)$$

Let's look at an example – a signal voltage $V_{in} = 0.05V$ is applied to the input of an amplifier with of gain $A_v = V_{out} / V_{in} = 20$

$$V_{out} = A_v \cdot V_{in} = 20 \cdot 0.05 V = 1V$$

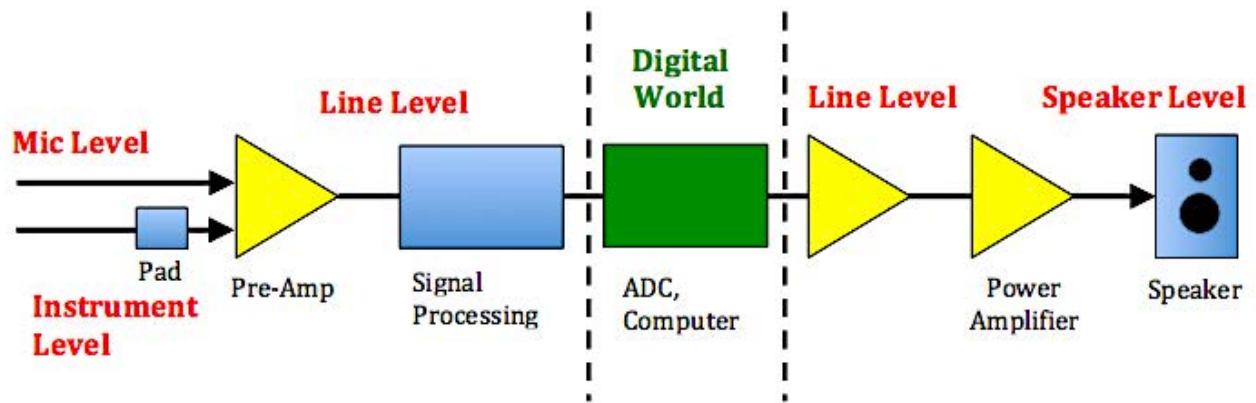
$$20 \log \left(\frac{V_{out}}{0.775 V} \right) = 20 \log \left(\frac{A_v V_{in}}{0.775 V} \right) = 20 \log \left(\frac{V_{in}}{0.775 V} \right) + 20 \log(A_v)$$

$$V_{out} (dBu) = V_{in} (dBu) + Gain (dB)$$

$$2.2 dBu = -23.8 dBu + 26.0 dB$$

So we have increased the signal voltage level by +26 dB, from -23.8 dBu up to +2.2 dBu.

The signal processing chain can be envisioned as follows:



The approximate range of signal voltage levels is provided in the following Table.

	Mic Level	Instrument Level	Line Level	Speaker Level
Voltage Range	1 - 200 mV	25 mv - 1 V	Commercial: 0.316 V Pro: 1.2 V	4 - 100 V
	-60 dBu to -12 dBu	-30 dBu to +2 dBu	Commercial: -10 dBV Pro: +4 dBu	+14 dBu to +42 dBu

For speaker levels, it is probably more meaningful to be talking about power. The power amplifier's impedance is matched to that of the speaker to provide **maximum power transfer** to the speaker. Under impedance-matched conditions, the power delivered to the speaker is,

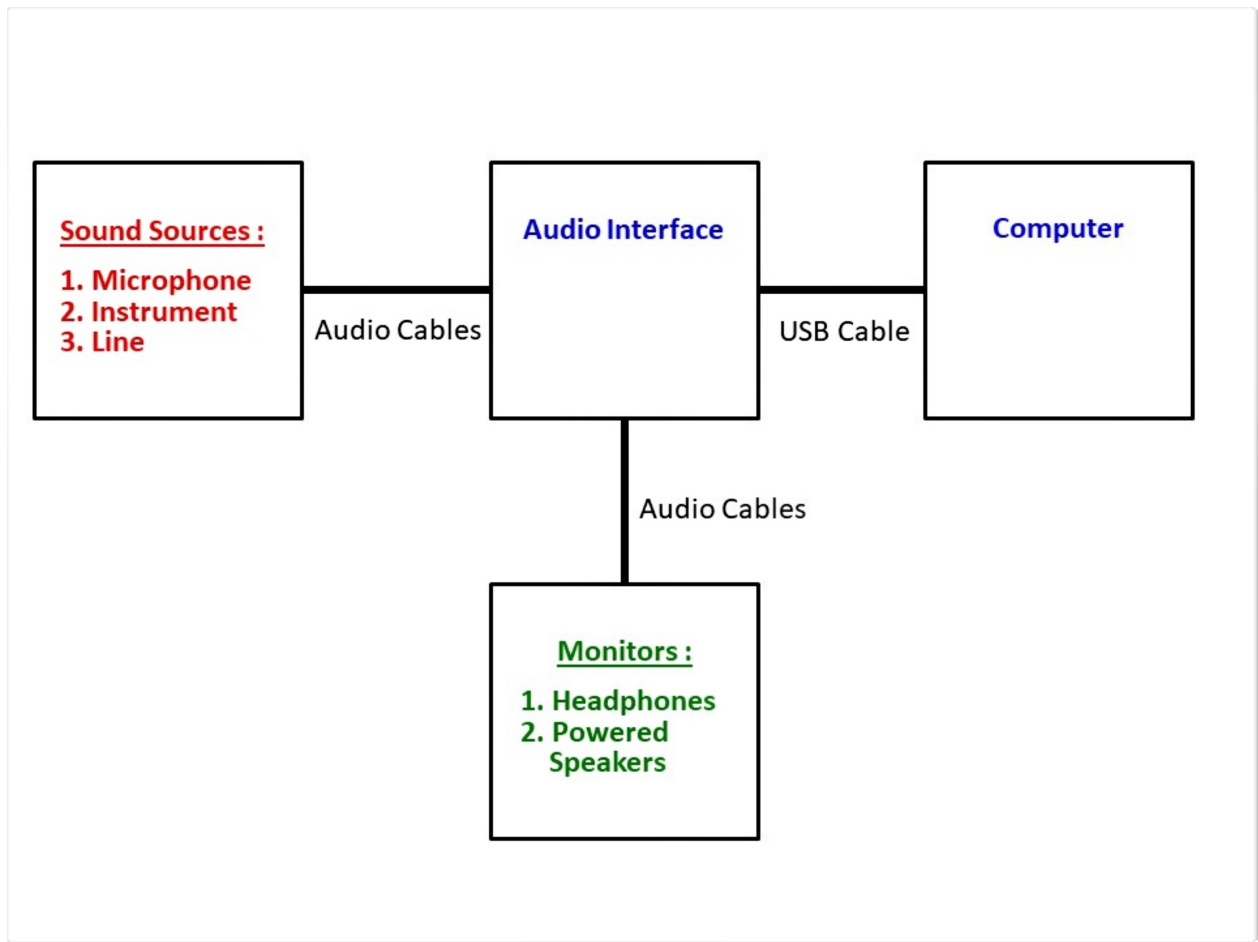
$$P = \frac{V^2}{R_L}$$

So for an average power $P = 150 \text{ W}$ delivered to an 8-Ohm speaker, the rms voltage $V = 35 \text{ V}$. The Thevenin equivalent source voltage of the power amplifier would be twice this value, $V_{TH} = 70 \text{ V}$.

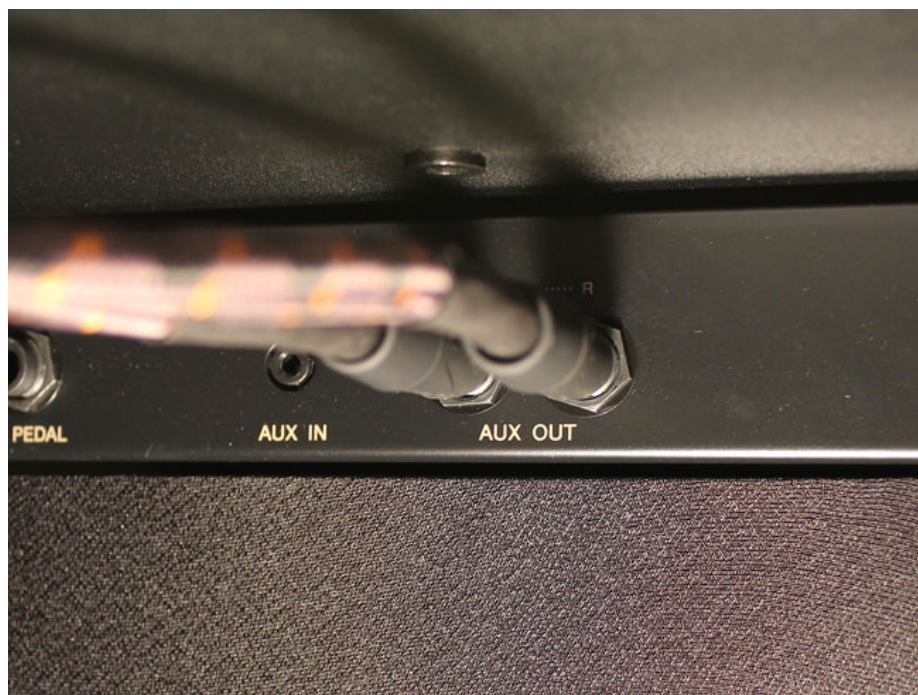
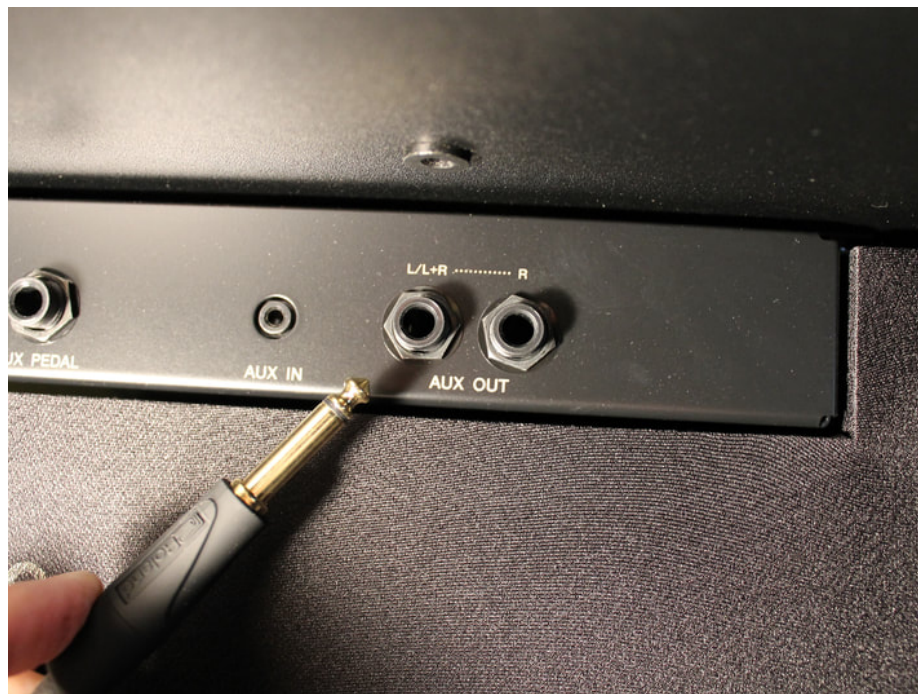
Audio Interface



Finally, let's hook up a sound source, like our digital piano, to the next major element in the recording chain, the audio interface. Here's a diagram of the primary building blocks of the home recording studio. You can see the central role played by the audio interface – it is at the intersection of the analog and digital worlds.



A pair of instrument cables (unbalanced, 1/4" TS plug connectors) are connected from the Left and Right Auxiliary Outputs of the piano to the two input "combo" jacks of the audio interface. The combo jacks accommodate connections from balanced XLR and 1/4" TRS cables, as well as unbalanced 1/4" TS cables.





There are three major operations that the audio interface is tasked with performing:

1. analog signal gain by a low-noise pre-amplifier
2. analog-to-digital and digital-to-analog conversion
3. high-speed data transfer to/from the computer.

All three of these operations are critical to making high-quality recordings. Choosing a good audio interface is therefore a very important matter in putting together your music studio. The prices of audio interface units range from budget to very expensive. I wanted to start with an inexpensive unit, learning to “walk” before “running”. I chose the **PreSonus Studio 24c Audio Interface**. I must say that this little unit is great -- its performance has been outstanding. The pre-amps are low noise and low distortion, and produce hi-fidelity sound. The analog/digital converters (ADC) are capable of sample rates of 44.1 KHz, 48 KHz, 88.2 KHz, 96 KHz, 176.4 KHz, and 192 KHz. The conversion resolution (bit depth) is 24 bit. The specified converter dynamic range is modest, at 108 dB, but this is still sufficient to record at levels allowing for 20 dB of headroom. We’ll talk more about the “digitization” of analog signals and what effect this has on the quality of sound in the next chapter. Lastly, the audio interface supports high-speed USB-C communication with the computer. The PreSonus Studio 24c is a class-compliant Core Audio device in macOS. No USB driver installation is necessary ! Audio interface units, such as the PreSonus Quantum, are now available that communicate with the computer via an ultra-low latency Thunderbolt-3 connection -- perhaps a gift to myself for next Christmas ? By the way, if you are thinking of purchasing equipment and have questions about audio interfaces (or any sound gear), I would recommend asking the folks at **Sweetwater**. Their very helpful staff are audio engineers and musicians, and are experts in the field.

The output connections on the rear of the PreSonus 24c audio interface are shown in the following photo. The stereo headphone jack (1/4” TRS stereo) is on the far right. The Main Output Left and Right jacks (1/4” TRS balanced) are connected via instrument cables to my left and right powered monitor speakers. The USB-C connection to my computer is on the far left. It should also be mentioned here that this audio interface unit is powered through the USB cable.



A final word – in the photo shown at the top of this post and above, you can see that the audio interface is sitting on a mat. The outer metal case of the unit makes contact to this grounded, anti-static mat. This is done in an over-abundance of caution to avoid any build-up of static charge that I may bring to the work table, especially in the winter time. I certainly don't want any crackling or pop sounds getting into the recording.

In the next chapter, we'll take a closer look at the "digitization" of the analog signal in the audio interface and what effects this has on the quality of the sound recording.

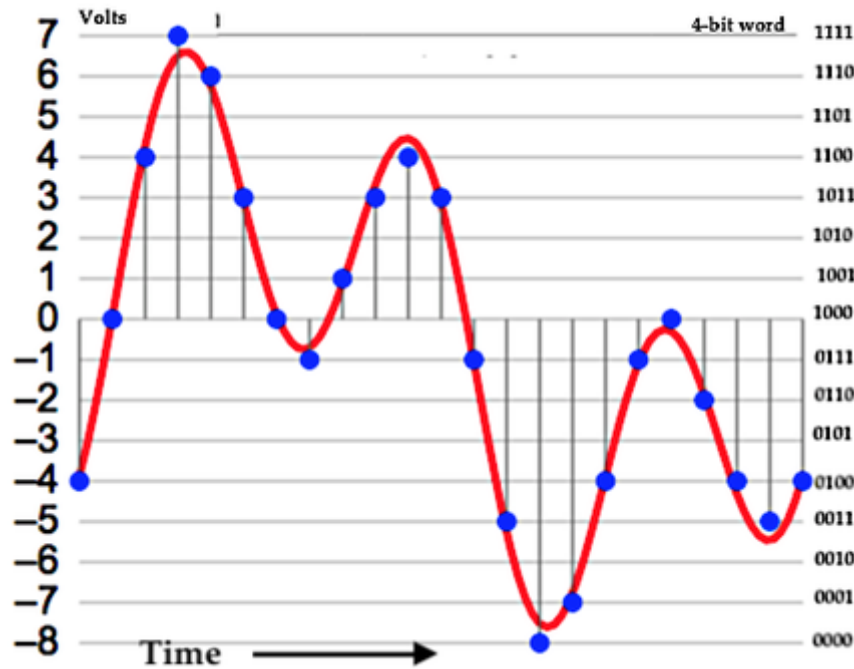
Digitization



The most important task of the audio interface is to convert an analog voltage signal to a digital stream of binary digits (bits) that represents the signal. This is the analog-to-digital conversion process in the **block diagram of the audio interface**. Its complementary process is the digital-to-analog conversion. This “digitization” process occurs at the interface between the analog world of the sound gear and the digital world of the computer. All of the “heavy lifting” of recording, editing, mixing, processing, and mastering the musical sound is done by software running on the computer – making this a computer-based recording system, or what is called an “inside-the-box” recording system.

The digital signal processing that takes place in the computer’s CPU is quite sophisticated and very interesting. It’s a subject that occupies university courses and curricula, and relies on some advanced discrete mathematics. Here, I will only touch on the subjects of sampling and quantization. More information, on an introductory level, can be found at the excellent learning resource **Digital Sound & Music: Concepts, Applications, and Science**.

“Digitization” of our analog signal consists of first sampling the voltage at a point in time, and then quantizing (assigning) that sample to the nearest regularly-spaced discrete voltage value. Each of these allowed discrete voltage values corresponds with a unique binary number. This process repeats at regularly-spaced intervals in time. It’s instructive to look at this process using the figure below.



Analog Signal Sampling and Quantization (modified from Auegg - Own work, CC BY-SA 3.0)

The continuous red-colored curve is the analog voltage signal $V(t)$. At regularly-spaced points in time, represented by the vertical grey lines, the signal is sampled, *i.e.*, the voltage is measured. The quantization process then “rounds” this value either up or down to the nearest allowed discrete voltage value, represented by the horizontal grey lines. In this example, there are 16 allowed voltage levels, separated by 1-volt increments. A binary number of 4 bits is used to label each of these voltage levels, as shown on the right side of the figure.

Clearly, sampling the analog waveform at a high rate and quantizing the voltage to many closely-spaced levels will yield a digitized waveform that is a “high fidelity” representation of the analog waveform. Let’s take a closer look at the sampling and quantization processes.

Sampling

How fast do we have to sample the analog waveform to get a digitized waveform that is capable of re-constructing the original sound with high fidelity? Fortunately, there exists the Nyquist Theorem of mathematics that gives us guidance in this regard. If we look at the frequency spectrum of a complex sound signal waveform (Fourier Transform of $V(t)$), we find that signal energy exists across a broad and nearly continuous range of frequencies that extends well beyond the human-audible range. The Nyquist Theorem states:

- Given a signal with maximum frequency component of f (Hz), a sampling rate of at least $2f$ is required in order to be able to re-construct the original analog waveform with full fidelity. The minimum acceptable sampling rate $2f$ is called the Nyquist rate.

- Given a sampling rate of f (number of samples per second), the highest-frequency component that can be correctly sampled is $f/2$. This highest-frequency component that can be correctly sampled is called the **Nyquist frequency**.

So, what happens if there exist frequency components in our signal that are higher than the Nyquist frequency $f/2$? Unfortunately, the spectral energies of these higher frequency components are "translated" down into the frequency spectrum below $f/2$ in a process called "**aliasing**". Not good -- this will lead to considerable distortion in our re-constructed analog waveform. To mitigate the effect of aliasing, we use an "anti-aliasing" low-pass filter on the signal (prior to the analog-to-digital converter) to remove the spectral energy at frequencies above $f/2$. Unfortunately, practical low-pass filters cannot have an infinitely sharp cut-off characteristic at the frequency $f/2$. To get around this fact, we actually sample at a slightly higher rate $f' > f$. Now, as the low-pass filter response rolls off above $f/2$, it can provide adequate attenuation of signal energy at frequencies at and above $f'/2$.

If 20 kHz is taken as the highest frequency that humans can hear (just an average value, depending very much on age!), then the Nyquist frequency $f/2 = 20$ kHz, and the minimum sampling rate should be $f = 40$ kHz (40,000 samples per second!). For CD-quality sound, the industry standard sampling rate is 44.1 kHz. Another industry standard rate is 48 kHz, which is commonly used for the audio part of video and DVD production.

Higher sampling rates are also readily available in most audio interface units. These sampling rates are multiples of two of the two base standard rates:

44.1 kHz, 88.2 kHz, 176.4 kHz

and

48 kHz, 96 kHz, 192 kHz

The 96 kHz sampling rate has been adopted as the *de facto* sampling rate for high-resolution recordings.

Why do we need to use these higher sampling rates? They produce large-sized data files and require substantial storage and media resources. But in their defense, it should be noted that the very color and timbre of the sounds of musical instruments come from the frequency spectral content that extends well above 20 kHz. While our ability to hear a pure sinewave tone may be dropping off gradually at higher frequencies, we may still retain a perception of a fuller color of a note sounded on a musical instrument if that high frequency content can be retained in the recorded audio waveform. There are varying opinions about this in the music recording and sound engineering community.

Quantization

When a sample is taken, the voltage amplitude of the signal at that moment in time is rounded to the nearest allowed voltage level which is represented by a unique binary number. The number of bits in this binary number is called the **bit depth n** , and determines the precision with which you can represent the sample amplitudes. The number of allowed voltage levels is given by:

$$\text{number of quantized levels} = 2^n$$

The resolution of the quantization process is dictated by the number of levels, and clearly this resolution increases rapidly with bit depth n . In the example figure above, the bit depth $n = 4$, so there are $2^4 = 16$ levels. The resolution is the voltage increment between the discrete levels, which is $15 \text{ V} / (16-1) = 1 \text{ volt}$ in the example figure. Now, look what happens when the bit depth $n = 16$. There are 65,536 levels, and the resolution would be

$$15 \text{ V} / (65,536-1) = 229 \text{ uV} (229 \times 10^{-6} \text{ V}) !$$

The difference between the original sample voltage and the quantized sample voltage that occurs from rounding is called **quantization error**. Quantization error is a form of correlated noise, since this error is an unwanted voltage added to the true amplitude voltage. Clearly, this error voltage is not a random process – it is strongly correlated with the signal. In this regard, the quantization error is better considered as **distortion**, as was done with correlated noise back in the posting on **Noise**. Because of this strong correlation with the true signal, the error voltage forms regular patterns in time that change in tandem with the original “correct” sound waveform. The error voltage itself constitutes an audio waveform, and this disagreeable sound could be very noticeable to a listener !

Another ‘source’ of distortion comes from **clipping**. Clipping occurs when the signal amplitude rises above the highest positive discrete level (the largest binary number, 1111 in the example figure) and falls below the lowest negative discrete level (the smallest binary number, 0000 in the example figure). In essence, the peaks of the sound waveform have been “clipped” off”. It’s possible that the error voltages during clipping are large enough to cause the total ruin of a sound recording. So recording at the proper audio levels is absolutely critical.

Keeping distortion to a minimum is made possible by using a sufficiently high bit depth n . The industry standard for CD-quality audio is $n = 16$. It is generally considered to be the minimum depth for professional audio production. A bit depth $n = 24$ is the current standard for high-definition audio applications, and is often used in conjunction with the 96 kHz sampling rate. In my recording sessions, I use the 96k/24 combination exclusively.

If higher bit depths are used, we run into a similar situation that was encountered with higher sampling rates – namely, does the resulting higher resolution go beyond the listener’s ability to discriminate it, and does it warrant the huge storage and media resources required.

A final word here on audio **dithering** – it’s a technique used to reduce quantization error introduced in the analog-to-digital conversion process or in the conversion of an existing data stream to a lower bit depth. This latter process occurs in practice when data processing at the 32-bit or 64-bit level in your computer’s digital audio workstation is interpolated back to the target bit depth at 16 bit or 24 bit. Interestingly, dithering works by adding a very small amount of random noise into the quantization

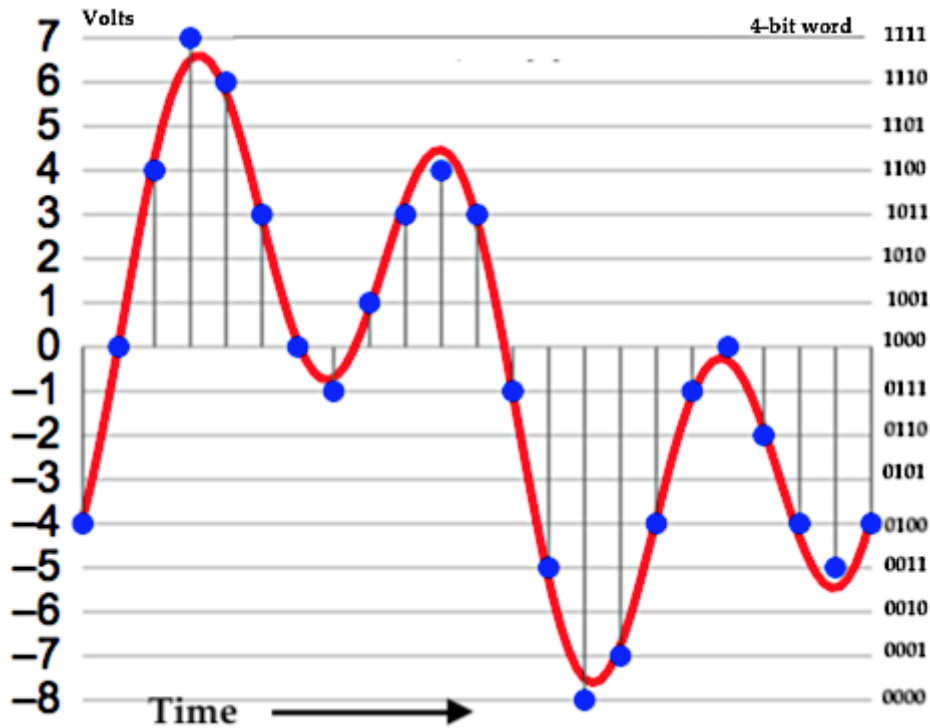
process, which breaks up the rounding error patterns that are correlated to the signal. In essence, a very small amount of random noise can reduce distortion caused by quantization error ! Nice !

Next time, we'll take a look at the topic of digital audio levels and dynamic range, which are intimately connected to bit depth.

A11. Dynamic Range



Previously, we looked at the various **audio signal levels** through the analog portion of the signal path. Here, we'll take a look at signal levels in the digital signal processing realm. The range of signal levels that can be recorded is limited by the finite number of quantized voltage levels available in the analog-to-digital conversion process. This finite number is determined fully by the bit depth used for the binary word that labels each of the quantized voltage levels. Let's refer back to the example used previously to describe the **digitization** of an analog signal.



$n = 4$ bit depth quantization (modified from Aquegg - Own work, CC BY-SA 3.0)

In the figure, the analog signal voltage varies in time, going almost equally above and below zero volts. Over a sufficiently large time window, the time-averaged voltage tends to zero for actual complex signal waveforms. So, we'll just focus on the positive voltage excursions above the zero-volts horizontal axis. In this example, a $n = 4$ bit word is used for the binary number labels. The most significant bit (MSB) in this word is the leftmost digit, which is "1" for positive voltages 0 to +7 volts. (The MSB is "0" for negative voltages.) If we ignore this MSB "sign bit", notice that the remaining binary word ranges from "000" to "111", yielding 8 discrete voltage levels for positive voltages. So the largest amplitude (loudest sound) voltage that can be recorded is given by:

$$|V_{max}| = \# \text{ intervals} \cdot \Delta V = (2^{n-1} - 1) \cdot \Delta V \quad (1)$$

where, ΔV is the uniform voltage step between quantized levels

The smallest amplitude (softest sound) voltage that can be recorded occurs halfway between the "000" and "001" levels, since this voltage will be rounded up and recorded at the "001" level. Any voltage less than this "halfway" voltage will be rounded down to 0 volts (silence), or equivalently, is masked by the quantization noise.

$$|V_{min}| = \frac{1}{2} \cdot \Delta V \quad (2)$$

Dynamic Range

Dynamic range is defined as the ratio between the loudest and quietest sounds that can be recorded. Given a bit depth of n , the dynamic range of a digital audio recording can be expressed, using $|V_{max}|$ and $|V_{min}|$ from above, as

$$\begin{aligned} \text{Dynamic Range} &= 20 \log \left[\frac{|V_{max}|}{|V_{min}|} \right] = 20 \log \left[\frac{(2^{n-1} - 1) \cdot \Delta V}{\frac{1}{2} \cdot \Delta V} \right] \\ &= 20 \log \left[2^n \left(1 - \frac{1}{2^{n-1}} \right) \right] = 20 \log[2^n] + 20 \log \left[\left(1 - \frac{1}{2^{n-1}} \right) \right] \\ &= 6.02 n - \frac{17.37}{2^n} \text{ (dB)} \end{aligned} \tag{3}$$

Let's calculate the dynamic range using the formula in (3) for the example in the figure above.

For bit depth $n = 4$, Dynamic Range = $24.08 - 1.09 = 22.99$ dB .

Now do a check by substituting the max and min voltages in (1) and (2) directly in (3), giving

$$\begin{aligned} |V_{max}| &= 7 \text{ V} \\ |V_{min}| &= \frac{1}{2} \cdot 1 \text{ V} = 0.5 \text{ V} \\ \text{Dynamic Range} &= 20 \log \left[\frac{|V_{max}|}{|V_{min}|} \right] = 20 \log \left[\frac{7 \text{ V}}{0.5 \text{ V}} \right] = 22.92 \text{ dB} \end{aligned}$$

which is in very good agreement with the previously calculated value.

In actual practice, bit depths are fairly large, $n > 10$. Therefore, we can drop the negligibly-small second term on the right-hand side of equation (3), and we get the well-known formula for dynamic range in digital audio,

$$\text{Dynamic Range} = 6.02 n \text{ (dB)} \tag{4}$$

Here are the theoretical values of dynamic range for the bit depths commonly used for recordings.

Bit Depth n	Dynamic Range (dB)
16	96.3
20	120.4
24	144.5

In the next chapter, we'll address how to 'measure' the signal amplitude level in the digital domain.

A12. Digital Signal Levels



For analog signals, we looked at the voltage amplitudes through the analog signal path in the previous chapter on **audio signal levels**. How do we ‘measure’ the signal amplitude level in the digital signal path? We know the highest allowed amplitude voltage in the quantization process is assigned the binary number whose digits are all “1’s”. We can define the amplitudes of the allowed voltage levels with reference to this “full scale” level as follows,

$$dBFS = 20 \log \left(\frac{i}{2^{n-1}} \right)$$

where i is an integer between 0 and 2^{n-1}

In our example of a bit depth $n = 4$, the highest level is $i = 2^3 = 8$. So this “full scale” digital signal level is specified by 0 dBFS, which is read “zero dB full scale”. The lowest (non-zero) level is $i = 1$. So this lowest digital signal level is specified by -18.1 dBFS.

For a more typical $n = 16$ bit depth digital system, we would have the following digital signal levels,

Integer i	Digital Signal Level (dBFS)
$2^{16-1} = 2^{15} = 32,768$	0 (full scale)
32,767	-0.00027
32,766	-0.00053
⋮	⋮
2	-84.29
1	-90.31
0	$-\infty$

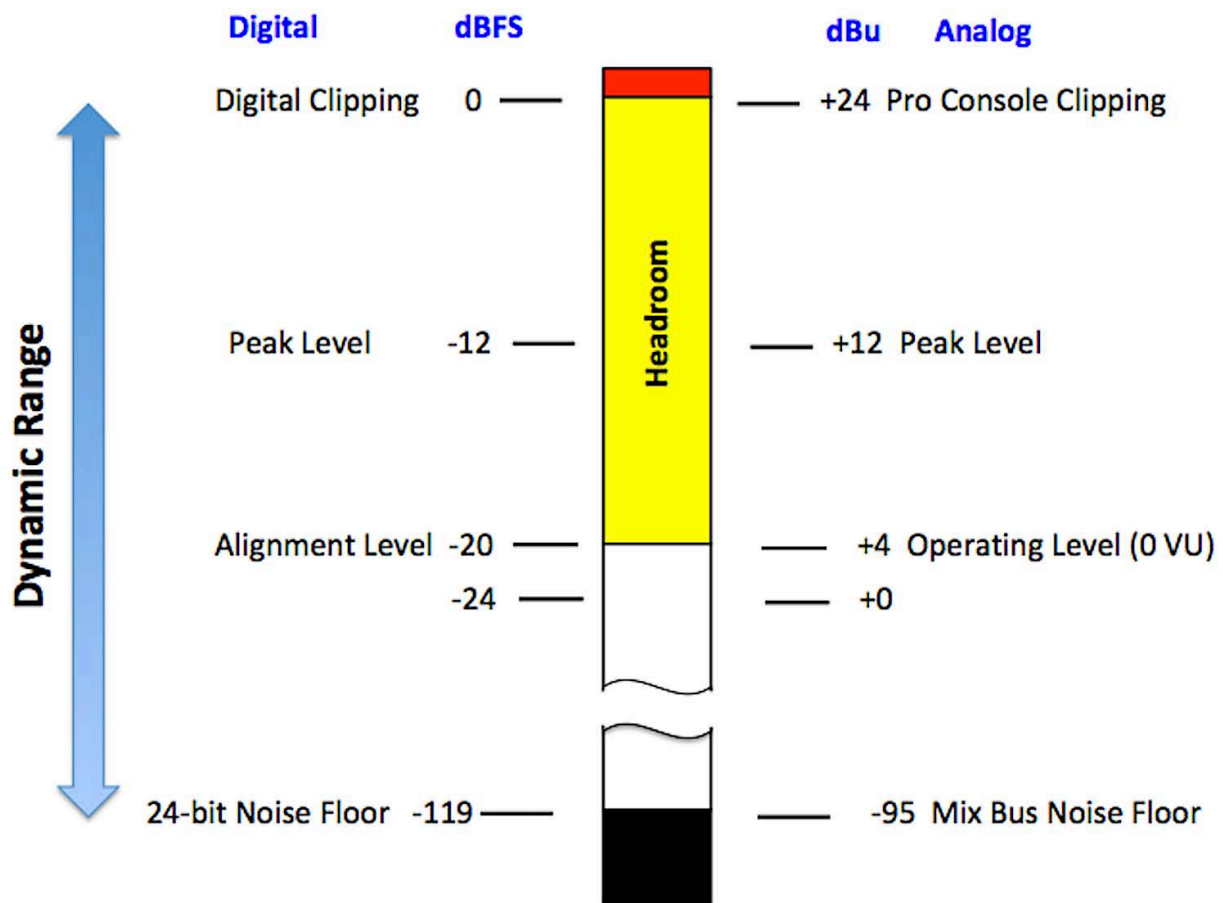
It should be kept in mind that dBFS is a logarithmic unit. In linear voltage units, the interval (step) between the huge number of allowed voltage levels is very small and uniform across the full dynamic range. In your digital audio workstation, you often see the signal input/output level meters expressed in dBFS units. An example of an input signal level meter in dBFS units is shown below.



Input Signal Level Meter dBFS

Gain Staging

An interesting question arises -- in the analog-to-digital converters used in the audio interfaces, what is the maximum voltage amplitude $|V_{max}|$ that is chosen to correspond to the digital full scale 0 dBFS ? There seems to be some guidance in this regard from various industry standards that align the digital signals levels (dBFS) to analog signal levels (dBu), such as in the following alignment,



This implies that the designers of the A/D converter use a value of $|V_{max}| = +24\text{dBu} = 12.28\text{ V}$. This may be true in high-end professional A/D units. But it's probably not the case in my budget audio interface that is powered via the USB bus. So different manufacturers may be designing to different $|V_{max}|$ values. The choice affects the high side of the dynamic range. On the low side of the dynamic range, the noise floor may be dictated by the existing noise present with the

signal from the analog circuitry, and not by the theoretical noise floor from quantization error. So the actual dynamic range available for recording music is certainly much less than the theoretical values for the different bit depths used in the digitization process, e.g., theoretical dynamic range = 144 dB for $n = 24$ bit depth. Actual dynamic range for bit depth $n = 24$ varies from 105 dB in inexpensive audio interfaces to 125 dB in top professional units. In the figure above, the dynamic range is shown to be 119 dB. The bottom line is that we would like to have a dynamic range available that well exceeds the dynamic range of comfortable human hearing, which is roughly 85-90 dB.

Managing the sound levels through the recording, mixing, and mastering processes is called "**gain staging**". The ultimate goal is to capture and recreate the full dynamic range needed for the type of music being recorded. There is A LOT written about this topic, some of which is just plain wrong. This article on **gain staging** from the folks at Sound On Sound provides an excellent overview. The key to setting proper input recording levels is maintaining sufficient "**headroom**" – the difference between the clipping point (0 dBFS) and the average sound level. Headroom is necessary to accommodate peaks in sound level as well as changes in level that will occur during signal processing and channel mixing, and ultimately in mastering the audio file. The current best practice is to allow for 20 dB of headroom. As seen in the figure above, using 20 dB of headroom still provides more than enough dynamic range for recording very quiet sounds. So when setting input levels, we should adjust the gain of the audio interface pre-amplifiers to yield average levels around -20 dBFS and keep anticipated peak levels around -12 dBFS.

The digital data created by the A/D converter in the audio interface is now transferred through the USB-C bus to the computer, the next major piece of equipment in the home music studio. We'll look at the computer hardware setup in the next chapter.

A13. The Computer



The "heart" of my home music studio is the computer. As such, it is important that the computer processor is high performance and the digital storage is large with fast input/output. Next to my digital piano, the computer hardware system will be an equally expensive element in the studio -- and rightfully so. Now is not the time to skimp, and it is important to plan for future recording and mixing needs that will require more computing power.

The universal question is whether to go Mac or PC. Since most of the important software tools and apps that are going to be used for recording and mixing are available in both macOS and Windows versions, the computer to choose boils down to personal preference. That said, however, there is a definite leaning toward Mac in the artistic community and in the recording industry. I chose the Mac, partially for this

reason and partially because the USB drivers to communicate with class-compliant Core Audio devices are already part of the macOS. This latter feature meant that my computer and my **audio interface** "talked" to each other automatically, with no effort on my part.

So, I spec'd the following computer (in the year 2000) to give me the processor speed and flash memory space that I wanted, without going too 'hog wild' and driving the costs too high :

Apple iMac 27" (shown in photo above)

macOS Catalina

3 GHz 6-core Intel Core i5 Processor

32 GB Random Access Memory (RAM)

256 GB Solid-State Hard Drive (SSD)

The macOS and recording/mixing software (digital audio workstation and signal-processing plug-ins) reside on the internal SSD. To store and retrieve the vast amounts of digital audio data, I keep the music files separately on an external hard drive with very fast input/output:

Patriot EVLVR 1 TB Thunderbolt 3 SSD

This solid-state PCI-express device allows high-speed data transfer at rates up to 40 Gb/s over the Thunderbolt 3 connection. The external SSD is shown below.

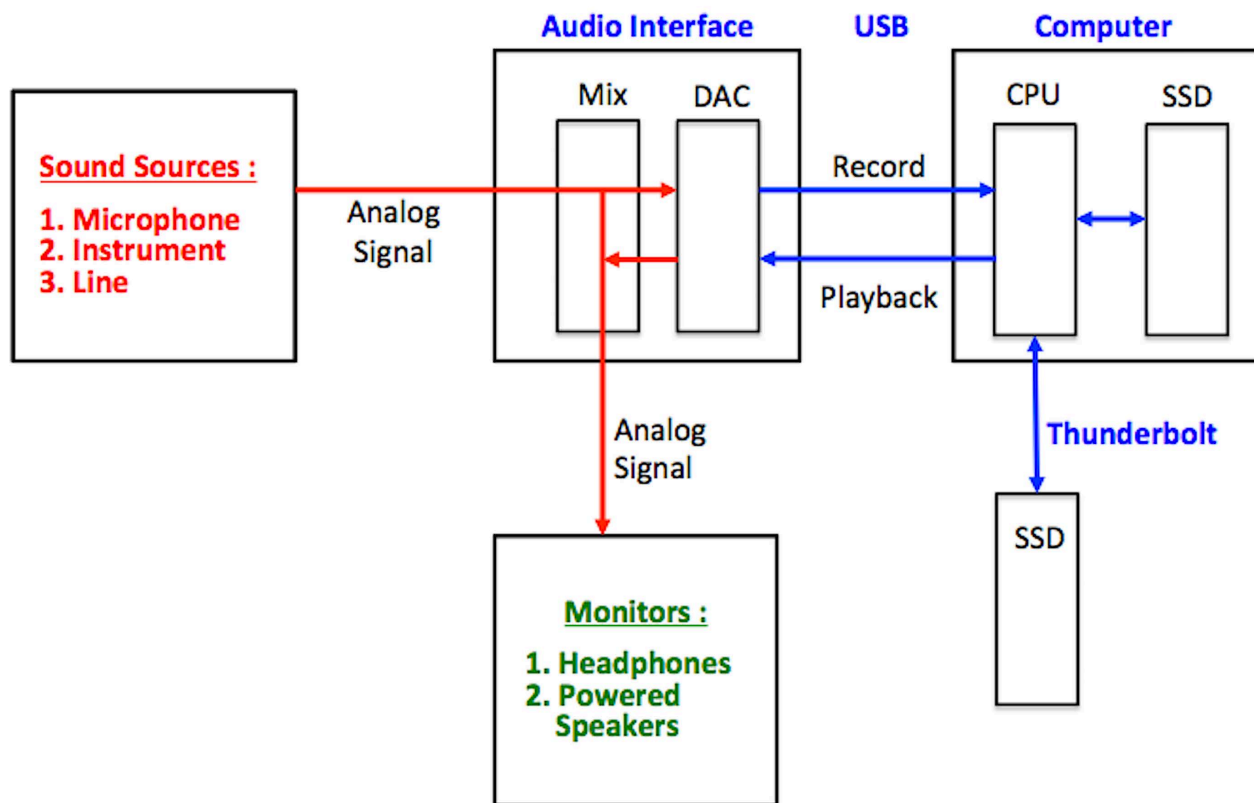


Having fast processor speed, plenty of RAM, and fast SSD devices will reduce the latency (delay) of the recording system and will protect against "drop outs" in the audio data. Latency will be the topic of the next chapter.

Finally, the hook-up of the computer to the audio interface is as simple as connecting the USB-C cable between them. Technically, this is a USB 3.1 connection with data transfer rates up to 10 Gb/s. (The "C" in USB-C refers to the miniature form factor of the connector itself.) It would be wonderful to have a Thunderbolt 3 connection between the computer and the audio interface, as this would reduce the latency of the system even more. Audio interfaces with Thunderbolt 3 connections are now available at reasonable costs (see [PreSonus](#)). The USB-C connection is shown in the photo below.



A14. Latency



Latency refers to the buildup of time delays, measured in milliseconds (ms), in digital audio signals as they pass through the hardware/software of the computer-based recording system. Looking at the digital signal flow in the block diagram above, we can see delays can be caused by the following processes:

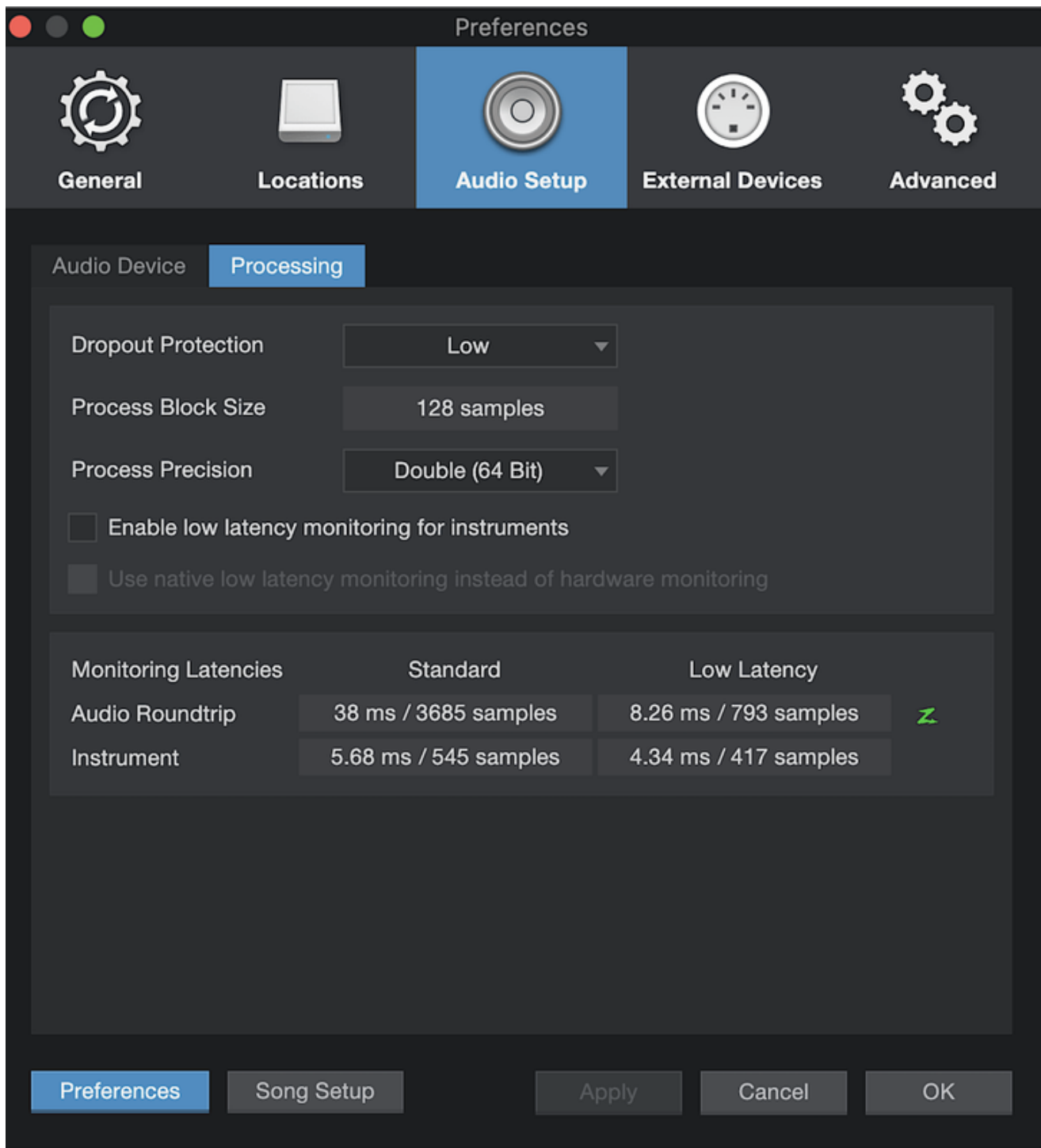
1. the analog-to-digital and digital-to-analog signal conversions in the DAC circuit of the audio interface,
2. the data transfer speeds of the USB or Thunderbolt buses between the audio interface and the computer,
3. signal processing speed of the digital audio workstation (DAW) software running on the computer CPU, and
4. data input/output speed of flash memory storage (solid-state drives (SSD)). The audio files are most likely kept on an external SSD, so a Thunderbolt-3 connection to the computer is necessary.

There are two activities associated with recording where latency can be a significant issue – monitoring and overdubbing.

Monitoring

When recording, you want to hear your performance as you play, as well as the performances of others if you're playing with a group. This is called monitoring. When monitoring the music being played through your computer's signal path, latency can be experienced as short delays between the time you play a note and the time you hear it on your headphones. If this delay is excessive, say, more than 8 – 10 ms, then it will be nearly impossible for you to play well on your own instrument, not to mention playing with a group.

The larger contributors to latency are numbers 2 and 3 in the list above. The availability of audio interfaces using Thunderbolt-3 (40 Gbps) connection to the computer has dramatically reduced latency. For computations made in the CPU for signal processing, there is a need for buffering data. By setting the buffer block size to a minimal amount (64 or 128 samples) in your DAW, we can reduce latency. A process block size of 128 samples is shown below for the **PreSonus Studio One DAW**.

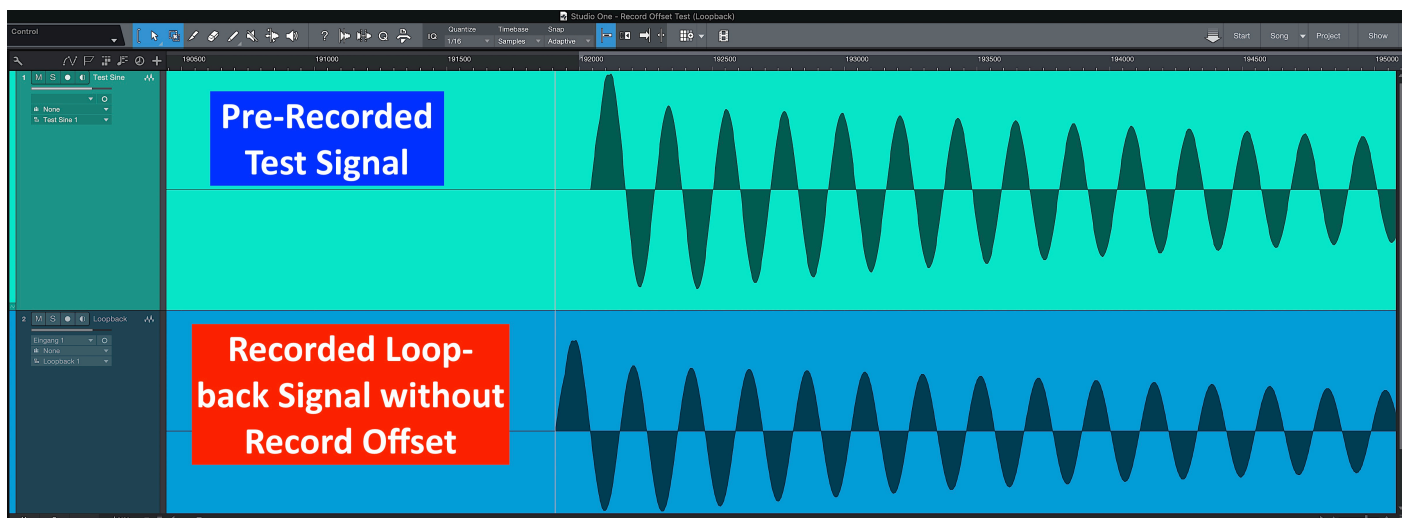


Using a small buffer block size is possible only if your CPU has the 'horsepower' to make the necessary computations quickly enough. Otherwise, data "drop outs" and system instability can often result. In the case where you must use larger buffer block sizes to have a stable system (causing unacceptable latency), most DAW software comes with a "low-latency" option for the monitoring signal path. In essence, the computer returns a portion of the digital audio data immediately back to the audio interface for monitoring purposes, without performing any significant signal processing. In the Studio One DAW example above, the round-trip audio monitoring latency is an unacceptable 38 ms. With the low-latency option engaged, the round-trip monitoring latency is reduced to a useable 8 ms.

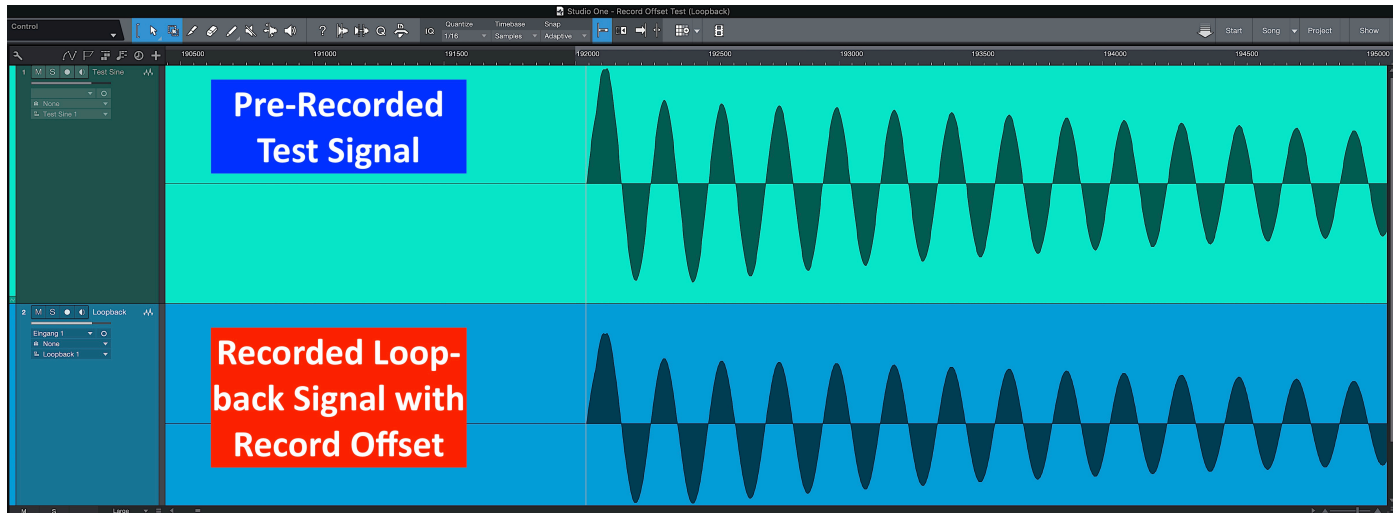
In keeping with the ‘philosophy’ of returning the monitor signal back to the performer as soon as possible, why not use the analog signal before even digitizing it ! Makes sense. And I believe this is what is routinely done in practice by many people, including me. Most quality audio interfaces have signal routing and monitoring mix capabilities -- this is shown by the red color arrows in the signal path block diagram at the top of this post. In this approach, called “hardware monitoring” , there is literally zero latency. You hear what you and your group are playing immediately in your headphones. Great !

Overdubbing

Overdubbing is the practice of listening to the playback of previously recorded tracks and recording an additional, separate track to the computer audio file that is “in synch” with the previously recorded tracks. The key word here is “in synch” -- timing is everything. Envision the signal flow using the block diagram at the top of this post. The playback signal originates from the audio file on an external SSD drive and travels all the way to the monitor output on the audio interface. Upon hearing this signal, you play right along with it, hopefully “in synch” . Your new signal at the audio interface input now travels all the way back to the external SSD drive and is recorded on a new track in the audio file. You can imagine that the round-trip latency will “place” the new track data somewhat delayed on the time line from the original playback track – *i.e.*, NOT in synch ! So, upon close inspection of the audio waveforms displayed by the DAW, I would expect the timing of the new track waveform to lag behind the original track waveforms. What I see, however, when I do this is that the new track waveform actually *PRECEDES* the original track waveform on the time line by some number of samples !! So my DAW is doing something that is not readily apparent, like trying to guess where to place the new track waveform properly on the time line. Only the software developers know for sure what is going on here. Fortunately, there exists a very short tutorial video from the PreSonus folks on performing a “loop back” test to ascertain how to align the two waveforms properly on the time line. A previously recorded transient on one track is played back to the monitor output of the audio interface and is then routed, via an audio cable, right back into the input of the audio interface to be recorded on a second track -- hence the name “loop back” test. Ideally, latency would be insignificant, and the two transient waveforms would be nearly aligned on the time line, *i.e.*, they would almost be “in synch” . But this screen shot shows the result :



The timing of the newly recorded transient *PRECEDES* the original transient by approximately 132 samples. By going deep inside the menus on the DAW, one can find, under the advanced audio settings, an input box labeled “Record Offset”. Here, you enter the 132 sample number. Rerun the loop back test. Lo and Behold -- the two transient waveforms line up nearly perfectly !! Amazing.

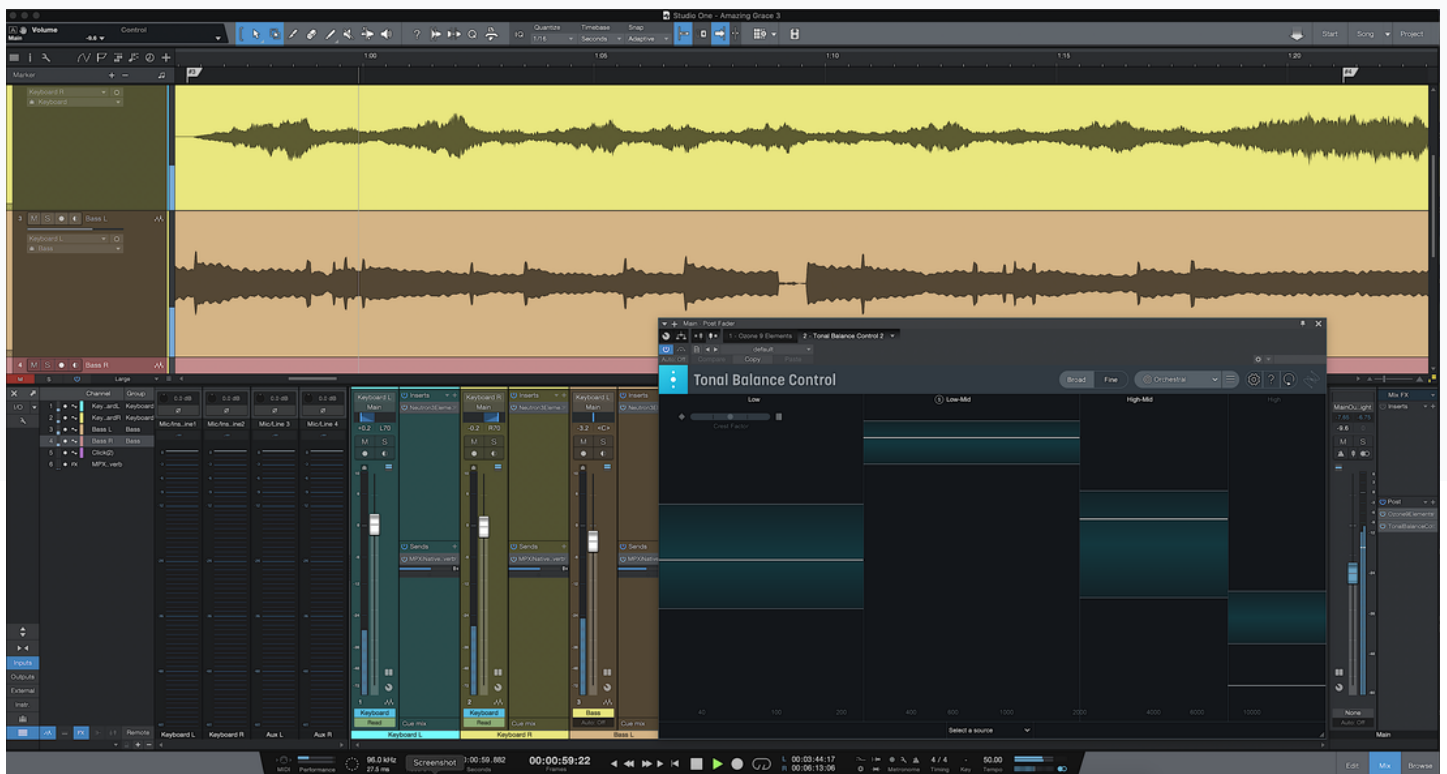


I am left wondering why such an important topic of aligning overdubbed tracks in the recording process is left to a relatively obscure tutorial and a menu item that is so deeply buried. It may be that the small offset is not readily detected aurally by a casual listener. But I am of the opinion that precision here is very important.

Note: The Record Offset value obtained from the loop back test is completely dependent on the settings for your recording -- sample rate, bit depth, computation precision, and buffer block size. Also, there are no FX plug-ins inserted in the playback channels. So if you change any of these, a loop back test would need to be run again to get the correct value for the Record Offset.

OK, in the next chapter, we'll take a look at the signal path in the digital audio workstation software for recording and mixing.

A15. Digital Audio Workstation: Signal Flow



PreSonus Studio One Digital Audio Workstation

An “inside-the-box” recording system relies on a sophisticated software application, called the **Digital Audio Workstation** (DAW), to perform the critical functions of music production: recording, editing, signal processing, mixing, and mastering. There are many DAW packages available to musicians, ranging from free “lite” versions (Audacity and Apple GarageBand) for the home hobbyist to expensive full-scale versions (Avid Pro Tools) for the professional studio engineer. Here’s a short list of the most widely used and highly rated DAWs:

PreSonus Studio One
Ableton Live
Image-Line FL Studio
Steinberg Cubase
Apple Logic Pro
Avid Pro Tools

The **PreSonus Studio One** is rapidly becoming a DAW that can rival the versatility and power of the Avid Pro Tools, but with a simpler learning curve and easier workflow. The latest release, Studio One 6, is bringing this DAW into the realm of industry-standard music production software. Working with Studio One is amazing – it is easy to learn, runs incredibly well (rarely, if ever, crashes), and has so many useful and powerful features. Essentially, anything I want to do, Studio One can do it. In fact, there is more capability in this DAW than I will ever know.

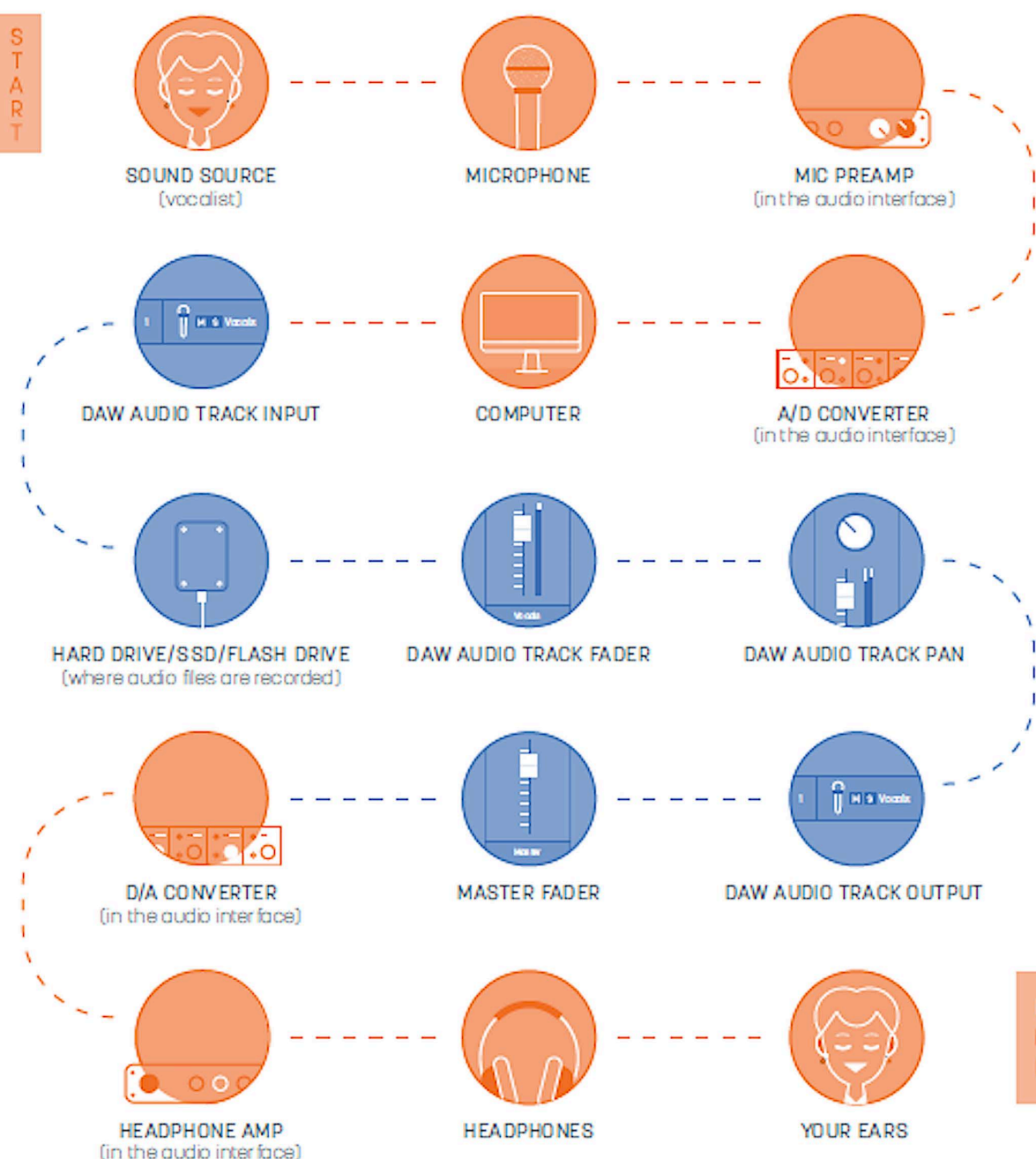
Signal Flow in a DAW

The signal flow in the DAW is important to understand in order to utilize the software properly. Unlike hardware components where you can visualize the signal flow in a system by the cables connecting the parts, signal flow in the software programming is not readily apparent to the user. Fortunately, a simple block diagram of the signal flow in a DAW is available for download from **iZotope**. The flow chart for recording is shown below.

SIGNAL FLOW CHART

BASICS FOR RECORDING

START



END

In the computer, the digital audio signal enters the DAW audio track input section, where input gain and phase shifts can be adjusted. It should be noted that the primary adjustment for recording level is usually done back at the analog pre-amplifiers in the **audio interface unit**. But the level of the digital signal can be "trimmed" at this point, just prior to the audio digital bits being recorded to the Hard Drive/SSD/Flash Drive. In the flow chart, the Hard Drive is the "final" destination of the signal in the recording process. All the remaining elements on the chart following the Hard Drive/SSD/Flash Drive are concerned with the monitoring process, and have no effect on the recorded audio files. For example, the DAW audio track fader does not affect the amplitude level of the recorded signal. This monitoring process during a recording session was discussed in the previous post on **latency**.

The signal flow chart for mixing is shown on the next page.

SIGNAL FLOW CHART

BASICS FOR MIXING

START



HARD DRIVE/SSD/FLASH DRIVE
(where audio files are played from)



INSERT EFFECTS



PRE-FADER SENDS



DAW AUDIO TRACK PAN



POST-FADER SENDS



DAW AUDIO TRACK FADER



DAW AUDIO TRACK OUTPUT



MASTER FADER



D/A CONVERTER
(in the audio interface)



YOUR EARS



HEADPHONES



HEADPHONE AMP
(in the audio interface)

END

Here, the signal flow mirrors the flow down the channel strip of a hardware mixing console. Playback of a track originates from the recording on the hard drive. The signal then passes through these basic elements of signal processing and mixing:

Insert Effects: Dynamics and Equalization Processing

Audio Track Fader: Level Control

Pre- and Post-Fader Sends: Time-Based Effects

Audio Track Pan: Stereo Image Processing

Main Output Bus: Stereo Mix

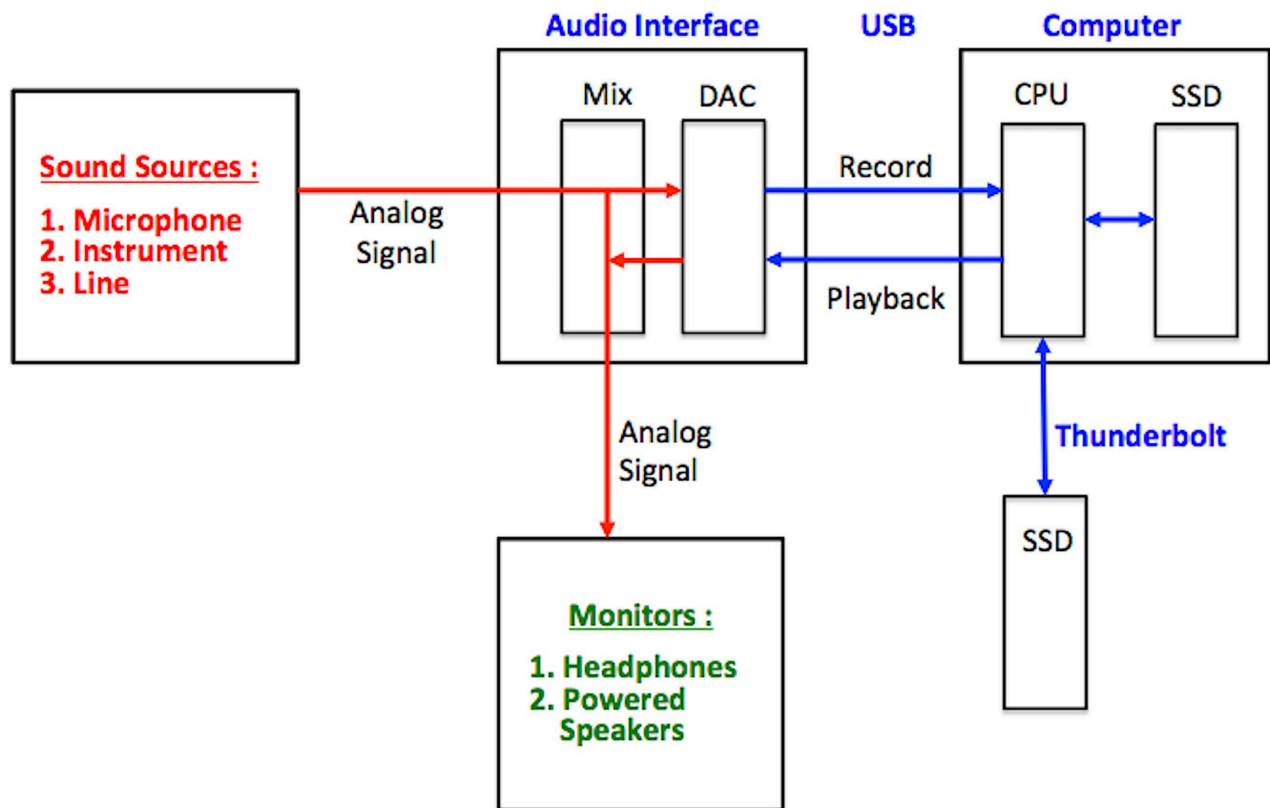
Master Fader: Main Output Level Control

Main Output Bus: Mastering Effects

The final mixed digital signal leaves the computer and returns to the audio interface unit, where it is converted to an analog signal.

The basic elements of signal processing and mixing listed above will be covered in Section C.

A16. Monitors



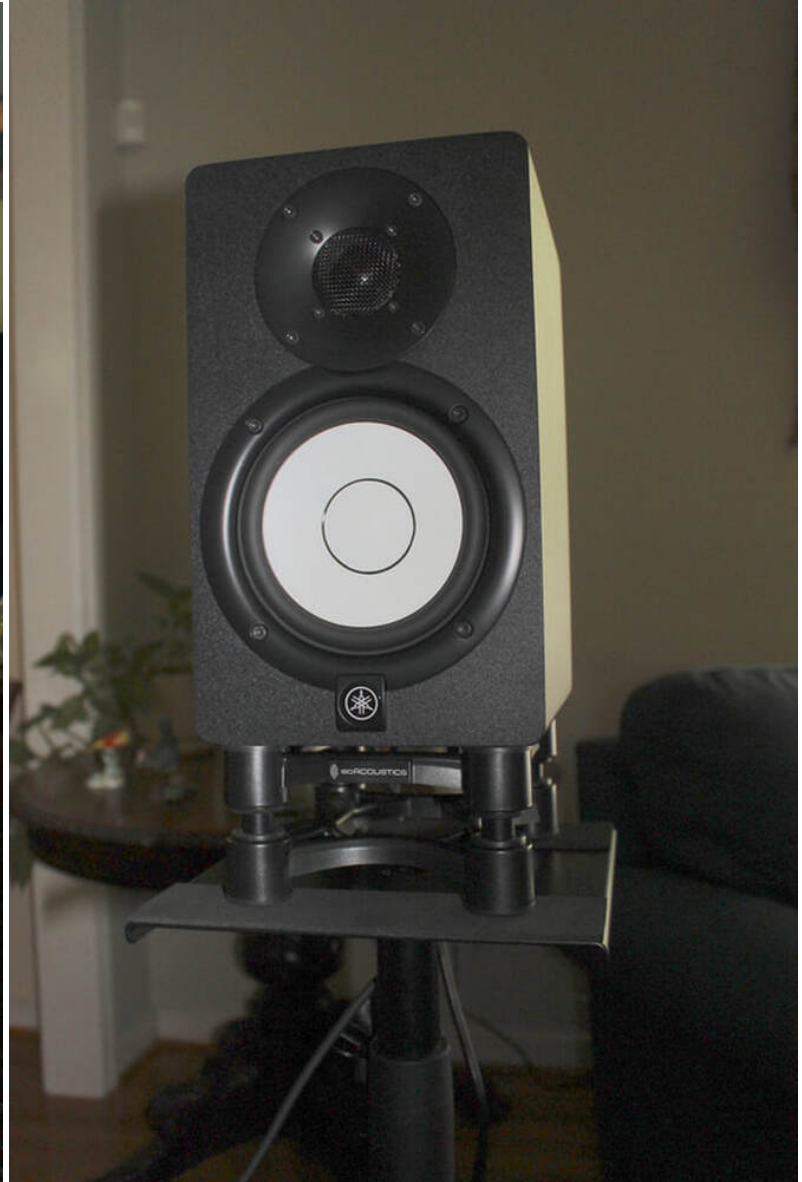
In the home music studio block diagram, I had reached the **computer** and the **digital audio workstation** software where the music recording, editing, mixing, and mastering occurs. I will return to talk much more about these music production processes in future chapters. But now it's time to complete the "hardware" in the studio with the **monitors** used to hear the playback of the music tracks.



Typically, there are two types of monitors used – headphones and powered speakers. These are connected to the analog signal outputs in the rear of the audio interface unit, as shown in the photos above. My Sennheiser HD 280 pro headphones are connected to the stereo TRS 1/4" jack. Yamaha HS5 powered speakers are connected to the left- and right-channel balanced TRS 1/4" jacks.



The balanced TRS 1/4" cable connection in the rear of the speaker enclosure is shown in the photo above. The speaker system incorporates a power amplifier to boost signal voltage **and** current to levels sufficient to drive the speakers. (This power amplification of line-level signals was discussed back in the chapter on **analog signal voltage levels**.) The speaker system consists of a two-way active crossover network that splits the line-level audio signal into high- and low-frequency bands. Each frequency band is then fed to its own specially-designed power amp, which in turn is used to drive the respective bass (woofer) and treble (tweeter) driver elements.

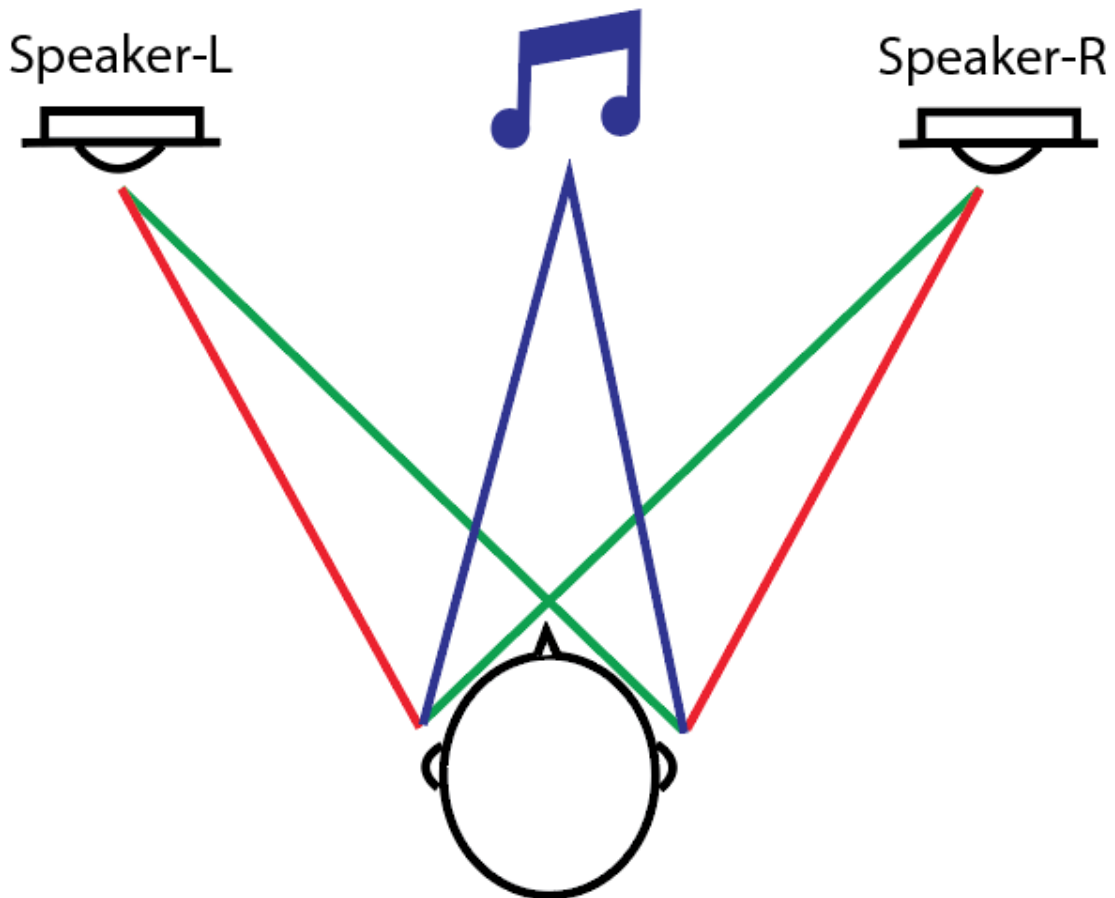


These speakers are designed to be “near field” monitors and to have a flat frequency response over the audible frequency range. “Near field” refers to the propagating sound pressure waves being fully formed in a short distance from the driver element. This is important because the home music studio space is typically small, and the direct “line of sight” distance between speaker and listener may only be on the order of 3 – 6 feet. The flat frequency response characteristic of the speaker is necessary so that the tonal balance of the music is not “colorized” by the speaker itself. As an example of why this is important, consider that a particular monitor over emphasizes the low-frequency content of the signal. The recording engineer will apply equalization in the mixdown to reduce the too-bass-heavy sound that he is hearing. But this actually leads to a tonal balance in the recording that is deficient in the low end. Not good ! The monitors are mounted on IsoAcoustics stands that isolate the speaker cabinets from the Gator Frameworks platforms on which the speakers are placed in the studio (see photos above). Natural resonances can occur between the speaker and the surface upon which the speaker rests. These resonances can add a distinct coloration to the sound from amplitude peaking at the resonant frequencies. The isolation stands act to decouple vibrations in the speaker cabinets from the supporting surface, thereby reducing the unwanted tonal coloration.

Finally, the sound waves emanating from the monitors interact in a complex way with the room acoustics and with the human hearing process itself. The public will listen to recordings over a very wide range of playback systems and speaker types, and under an infinite number of listening conditions. It is only possible to record and mix music that sounds “just right” on your sound gear, in your studio space. But by having quality sound gear, conditioned room acoustics, and a good ear, you can make recordings that will sound great to a majority of listeners.

In the next two chapters, I’ll take a look at the science, and art, of setting up the monitors in the studio space.

A17. Sound Source Localization



In the field of psychoacoustics, the ability of the human auditory system (two ears and brain !) to identify the direction from which a sound is emanating is called sound source localization. Humans can locate this sound source in space with extreme precision – within 2 degrees in the horizontal plane. This remarkable feat is accomplished by the brain's ability to process the binaural signal coming from the ears.

Neuroscientists believe that sound localization in the horizontal plane relies on two “cues”: the sound amplitude (loudness) difference between the two ears (inter-aural level difference ILD), and the time difference (delay) of sound reaching each ear (inter-aural time difference ITD). The brain uses both cues to localize sound sources. While variations of either can evoke a perceived

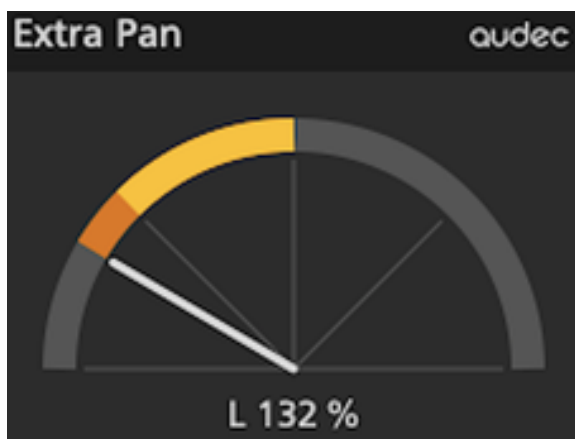
direction change, ITDs and ILDs are not completely independent of each other. There exists a natural cohesive spatial relationship between ITDs and ILDs. For example, suppose you have a single speaker off to your left side. The sound (wave) coming from the speaker would reach your left ear sooner and be louder than the sound that reaches your right ear. Your brain compares these differences and interprets where the sound is coming from.

Our stereo music systems employ two sources of sound: a left speaker and a right speaker. The stereo field is the spatial “width” between the two speakers. For a listener sitting in front of and centered between the speakers, as shown in the figure above, sound emanating from the left speaker reaches your left ear (“direct” red line) and your right ear (“crosstalk” green line). Same thing for the right speaker and your two ears. There is a summation of the direct and crosstalk signals at each ear. Our perception of a single source of sound, a “phantom image”, coming from somewhere between the speakers will be based on the inter-aural level difference (ILD). For example, if identical signals at the same volume level are played from the two speakers, the listener will perceive that the source of the sound is located exactly midway between the speakers, as there is no level difference between the ears. This situation is indicated in the figure above by the centered source (blue musical notes).

This is the concept of “panning” (placing) a musical track in the stereo field. By adjusting the relative volume levels of the track signal sent to the Left and Right channels of the Main stereo output of the mixer console, we can move the location of that track’s sound in a continuous fashion from Left speaker all the way to Right speaker. As an example, a “hard-panned-Left” track will have volume level in the Left speaker only – there is zero volume level in the Right speaker. Consequently, the image location of that instrument in the stereo field is at the left speaker.

At this point in the discussion, I want to mention briefly the idea of “pushing” the sound image beyond the speakers, *i.e.*, perceiving the sound location as being further out on the sides, past the locations of the speakers. You may want to do this to “widen” the stereo field. As will be discussed in the chapter on Mid-Side processing, this can be done ! For example, say a signal is already panned completely to the Left speaker (so there is zero signal output in the Right speaker.) A small, phase-inverted version of this signal can be placed in the Right speaker. Upon summation of the waves coming from both speakers at both ears, there is a differing amount of destructive interference caused by the phase-inverted signal at the two ear locations. This in turn leads to a greater inter-aural level difference (ILD) and the perception that the sound is coming from a direction even further to the left of the left speaker !

In addition to using **Mid-Side processing** in an Imager plug-in, there are specialized plug-ins that use this phase-inverted signal technique to widen the stereo field, such as the **Audec Extra Pan** plug-in.



Pushing the sound image beyond the left speaker (L 100%)

This “outside-the-speakers” technique, however attractive, must be used with caution, since playing around with phasing can lead to considerable problems, such as sound drop-outs. So, heed the warning.

In the next chapter, I'll discuss the room acoustics of your home music studio and the placement of your powered monitors.

A18. Acoustics



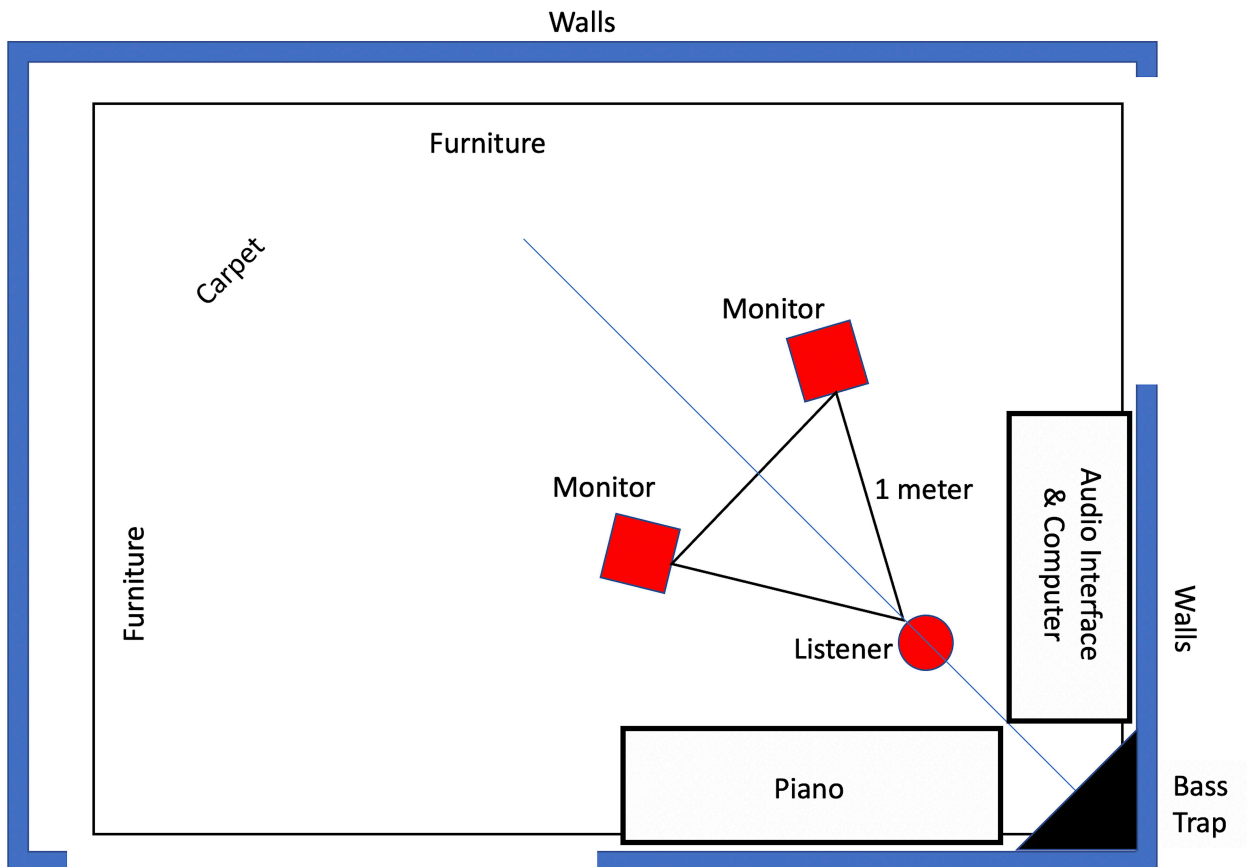
In the last chapter, we discussed the **monitors** used to listen to the recorded music. The sound characteristics of these monitors play a key role in achieving good sound mixing and tonal balance in the final tracks. But the sound we hear coming from the monitors is greatly influenced by the acoustics of the room. So it is vitally important to set up the home studio properly so that we can hear the “true” musical sound. We don’t want to be “fooled” by acoustic artifacts of the space.

A professional studio will design and build the space using standard acoustic guidelines. A home studio is greatly constrained by the fact that selection of available room or space is limited and that there is only a modest budget to fix up the space with appropriate acoustic treatments. Nevertheless, certain ground rules of acoustical physics should be followed in order to create a good listening

environment. Some of the more important considerations are listed below.

1. Symmetry

Acoustic imaging (ability to discriminate placement and balance in the stereo field) is best when the listener, speakers, walls and other acoustical boundaries are symmetrically centered about the listener's position. This is particularly challenging if you don't have available an entire room (like a bedroom) to use for the studio. I was constrained to use just the corner of our living room. After a couple iterations, I settled on the following layout of the sound gear.



The two speakers and the listener are at the points of an equilateral triangle – each side is about 1 meter in length. Hence the left and right stereo channels are equidistant from my ears, and I am positioned at the center of the panning image. My monitors are placed on stands in the interior of the room, with the tweeter drivers at the height of my ears while I am sitting in my chair. The tip of the triangle at my position is pointed symmetrically (along the 45-degree angle line shown) into the corner of the living room. Any first-order reflections of the sound waves from each speaker off its nearest wall are also equidistant to my ears, so the time delay is also equal. Ideally, I would have sound absorber panels on the walls at these "mirror points", but there is at least some "stuff" (like the piano, sound gear table and wall mountings) in the pathways, reducing the specular reflection of these waves.

2. Standing Wave Effects

Sound waves reflect quite well off flat solid surfaces, like room walls, and the resulting interference pattern of incident and reflected waves creates “hot” and “cold” spots, playing havoc with the frequency balance in the room. (*i.e.*, the space does not exhibit a relatively flat frequency response over the audio spectrum at all positions in the room.) Parallel walls in rectangular rooms are particularly bad in this regard since the size of typical rooms are on the order of multiples of half-wavelengths of the audio signals. As an example, the frequencies that could establish noticeable resonant modes in a room with a 14 ft. separation between parallel walls are given by,

$$f = n \frac{v}{2L} = n \frac{1130 \text{ ft./sec}}{2 * 14 \text{ ft.}} \cong n * 40 \text{ Hz}$$

n multiple	Frequency f (Hz)
1	40
2	80
3	120
4	160

At these and higher frequencies, you would detect amplitude peaks and dips (anti-nodes and nodes) in the sound as you moved around the space.

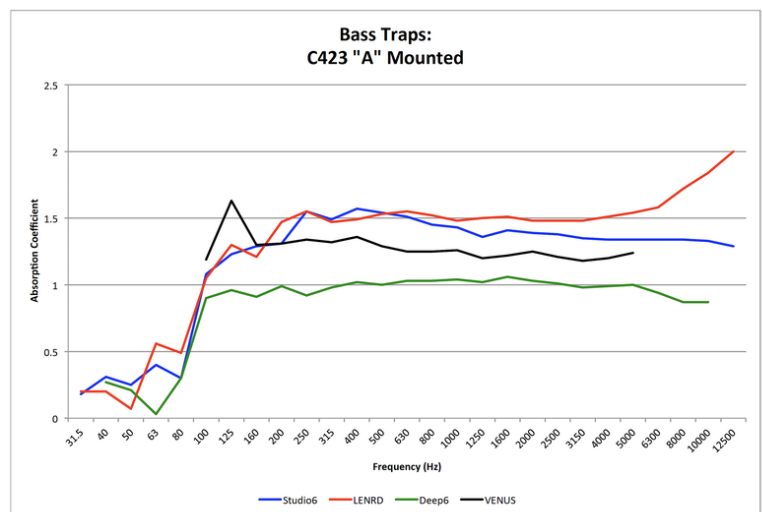
In order to avoid this unwanted effect, you want to “break up” the unobstructed path between the two parallel surfaces, so that the wave fronts scatter over a large solid angle. This can be done by inserting complex surfaces throughout the room, or, as is common in studios, by mounting diffusing panels on the walls so that reflecting waves disperse in random directions. In my setup, the sound waves propagate past my listening position into complex surfaces with irregular geometries. This serves to diffuse and absorb the waves, and helps to avoid the formation of standing waves.

3. Bass Buildup Effect

Whenever possible, speaker enclosures should be placed at least 2 feet away from the nearest wall and/or corner. This helps reduce the bass buildup that acoustically occurs at boundary and 90-degree corner locations due to wave reflection. (The boundary condition for the sound wave pressure is an anti-node.)

This effect was a problem for my setup when I initially had my monitors close to the walls in the corner of my living room. And the effect was exacerbated by the fact that the speaker bass reflex ports (the open apertures in the speaker enclosure) are in the rear of my Yamaha monitors ! Fortunately, my final setup put my monitors on stands, out in the middle of the room, as shown in the drawing above.

Since I am pointing my speakers into a corner, I did feel obligated to place “bass traps” in the corner to reduce the bass buildup effect there. I used two **Auralex LENRD bass traps** to absorb the high sound pressure. This was a fairly inexpensive acoustic treatment to apply.



4. Reverberation

“Reverb” is the persistence of a sound signal after the direct “line of sight” signal has ceased. This time persistence is caused by the large number of multiple reflections that come back to the listener after traveling over many different path lengths with their associated time delays. Reverberation time is characteristic of the “depth” and nature of the space. A music performance in a small recital room sounds vastly different from the same performance in a stone cathedral ! This room ambience is an integral part of the listener’s experience of a live performance, and needs to be captured in the recording. On the other hand, in the music studio control room, where you are doing the mixing and sound engineering, you really don’t want the control room’s reverberation characteristics to obscure the sound that you are carefully listening to. But it would be similarly undesirable to mix in a completely “dead” anechoic chamber. Certainly, the public will listen to the recording in a somewhat “live” room. I feel that my living room, with a modest amount of furniture, wall treatments and carpeting, has a warm subtle ambience suited for mixing using my monitors.



A19. Un-interruptible Power Supply (UPS)



APC Back-UPS Pro BX1500M

I'm concluding Section A with a very important chapter about the power quality supplied to the critical electronic circuits in the recording chain. All circuits that create, process, record, and play back signal data require electrical power obtained from the wall outlets in your home. This AC power needs to be properly grounded, filtered, regulated, and uninterrupted. The most important equipment you "plug" into your home's wall outlets includes electronic instruments (digital piano), the audio interface (with its pre-amplifiers and ADC converters), and of course, the heart of your recording system, the computer.

The importance of having quality power became very apparent to me when a very fast power transient in the line circuit into which my computer is plugged caused the computer to flick off. Certainly, the loss of data can be catastrophic. But I also wondered about 'noisy' and unregulated power being supplied to amplifier and other signal processing electronic circuits in the recording chain. Could there be any noise or distortion effect introduced into the voltage signal ? In any case, it was time to invest in an uninterruptible power supply (UPS) unit that would provide uninterrupted and conditioned power to my electronic circuits.

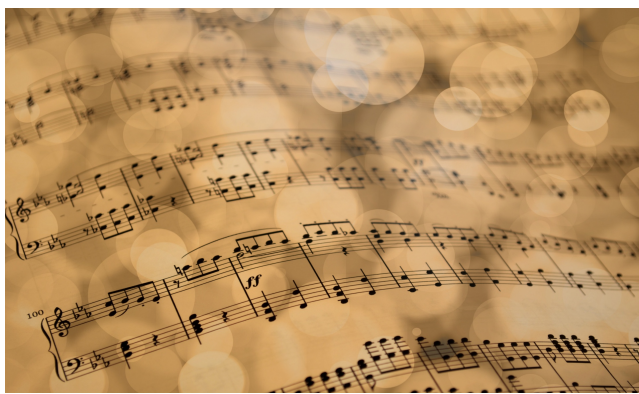
A good UPS unit protects your electronic devices from harmful power surges, spikes, lightning, and outages. It also provides noise filtering and automatic voltage regulation that instantly corrects voltage fluctuations. I purchased the **APC Back-UPS Pro BX1500M** unit, shown in the photo above, to use in my home music studio. This unit has the following specifications:

Output power capacity (volt-amps):	1500VA
Maximum Load:	900 Watts
Utility AC Voltage Range / Frequency :	88 to 139Vac, 60Hz +/- 1Hz
Automatic Voltage Regulation (Vac):	(88-107) + 11.2 %
Output Voltage / Frequency (On battery):	115 Vac +/- 8%, 60Hz +/- 1Hz
On-battery Waveshape:	Stepped Approximation to Sine Wave
Transfer Time:	8 ms (typical), 10 ms (maximum)
Battery Type:	24V sealed, lead-acid
Surge Protection Energy Rating:	789 Joules
Filtering:	Full-time multi-pole noise filtering 5% of IEEE surge let-through zero clamping response time: instantaneous

The cost of this UPS unit was \$165 in year 2020. Although this may seem rather "pricey," I feel it is a worthwhile expense to protect your home music studio equipment.

B. Recording

B1. Recording and Editing



Largo - George F. Handel

click on music score to play

My first recording, along with my first attempts at editing, in my home studio was a portion of a piano transcription of Handel's Largo from his opera Xerxes.

When the recording "red light" comes on, it's only natural to feel 'nervous' , just like you would feel performing the music in front of a live audience. You're thinking: "This is it. What I record into the digital audio file is permanent. I've got to "get it" just the way I want, so that listeners will hear and enjoy a good musical performance."

This is why it is necessary to have done all the hard work of practicing the piece to a sufficient level of technical mastery beforehand. And then, take a lot of time (perhaps years) to allow your playing to develop and mature the musical ideas you wish to convey. Now, you can concentrate on the interpretation and musicality of the piece while recording the performance. A good warm-up on your instrument before hitting the record button is also highly recommended !

It is unreasonable, however, to expect to be able to "lay down" your very best performance in a single complete "take" . It is very comforting to know that you can, and will, use some editing of the recording, during the recording session as well as afterwards, to put together a performance you are happy with. I'll review some of the more commonly used editing techniques a bit later. But I want to give you now some excellent thoughts about the philosophy of editing expressed by the pianist **Paul Cantrell** on his website:

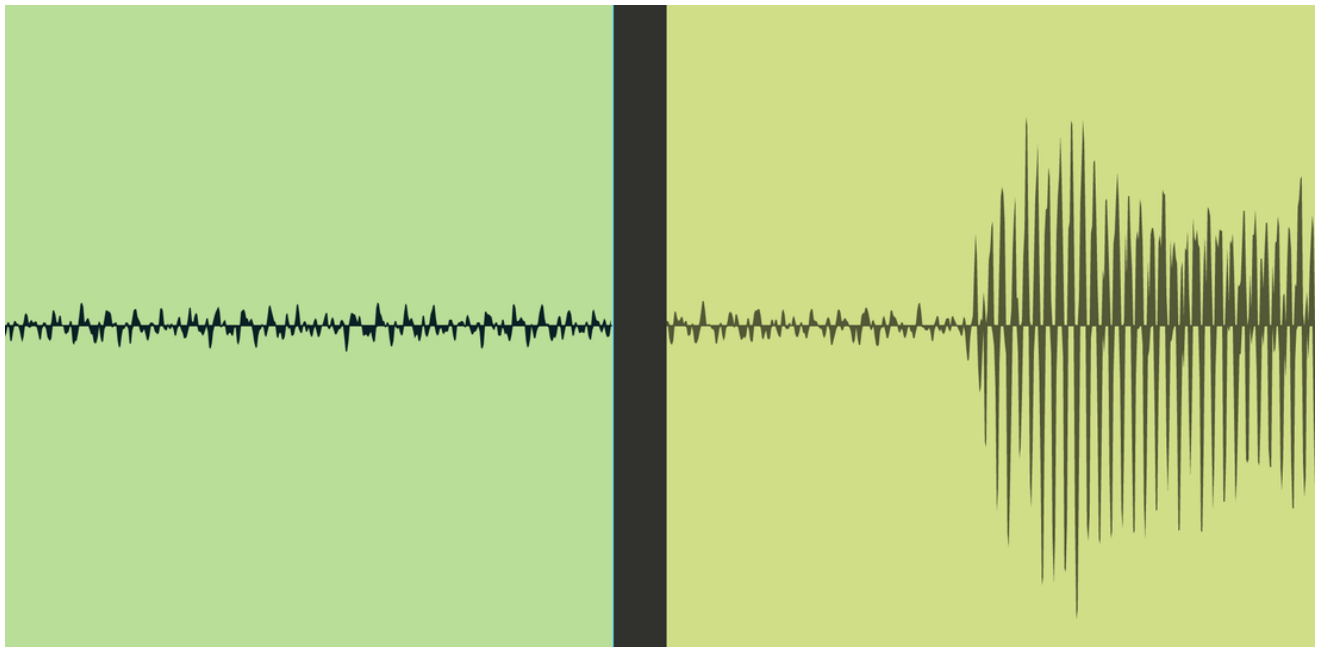
An aside on the philosophy of splicing: I'm skeptical of classical recordings with hundreds of splices that are pieced together entirely in the studio — it works well when the studio is part of the compositional process, but when the purpose of such splicing is obsessively perfectionist correction of live playing, it is dangerous. It can lead to a clinically perfect but spiritless recording without the organic expressiveness that makes music magical. *However*, neither am I in the camp that decries splicing as some kind of hoodwink or moral failing; recordings are recordings, not live performances, and a musician's job is to work their medium's full potential to produce the best possible experience. So, when I splice, I try to strike a balance between correcting really conspicuous mistakes that disturb the flow of the music, and preserving that flow in its natural form. In short: splices are artistic decisions, and must be treated as an aspect of musical performance, not as cosmetic surgery. End of aside.

Editing occurs during the recording process as well as in the post production phase. In Classical music, especially from the Romantic and subsequent periods, there are continuously changing tempi and a wide dynamic range in a piece – the music flows with rhythmic and expressive freedom. In contrast, multi-track recordings of contemporary music typically use a "click track" that keeps the time, like a metronome, and allows for overdubbing parts. (Click tracks can be programmed to have tempo changes within the song, but not in the "rubato" sense.) The use of a click track also greatly facilitates editing a recording through the techniques of "comping" (compositing), "punching in" (replacing a few notes or a phrase) and slip editing (aligning notes between tracks). In Classical

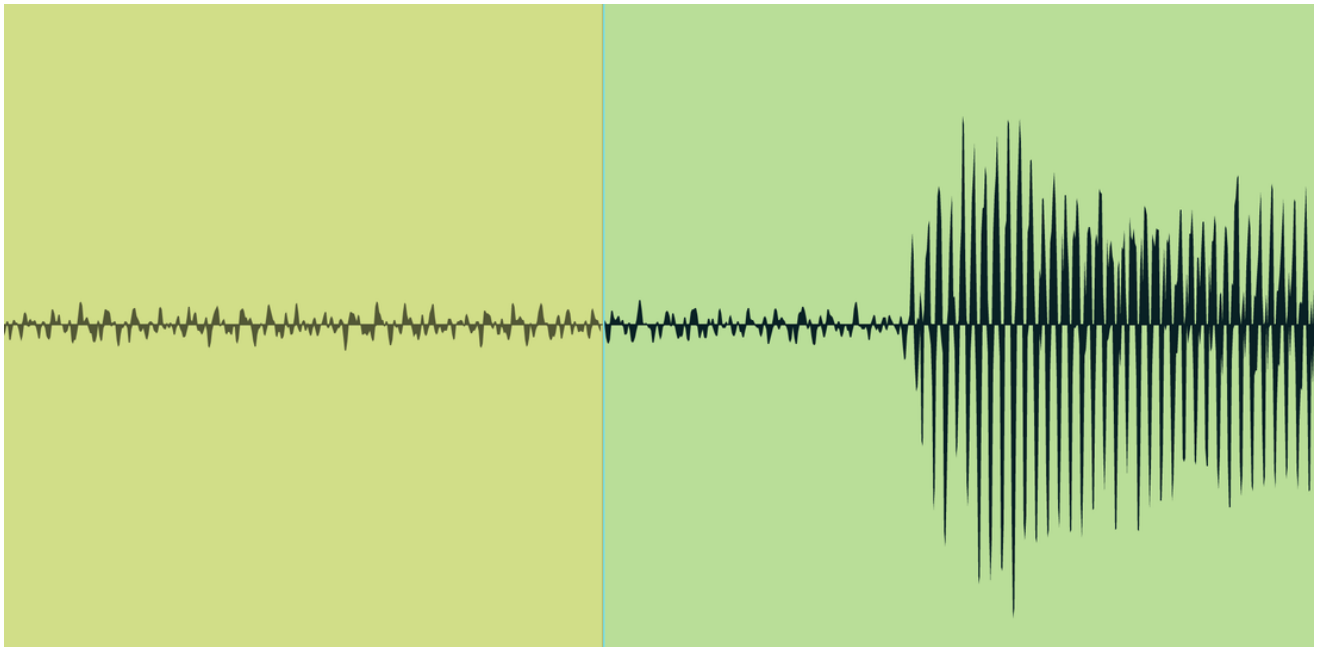
music, editing a recording can be done using a form of comping – splicing together good takes of sections of the piece. But given the natural and emotional flow of the music, there exist only a few points in a piece where two sections can be joined successfully. Two such points are discussed below. The waveforms shown are from the recording of the Handel Largo.

1. Crossfading over silence

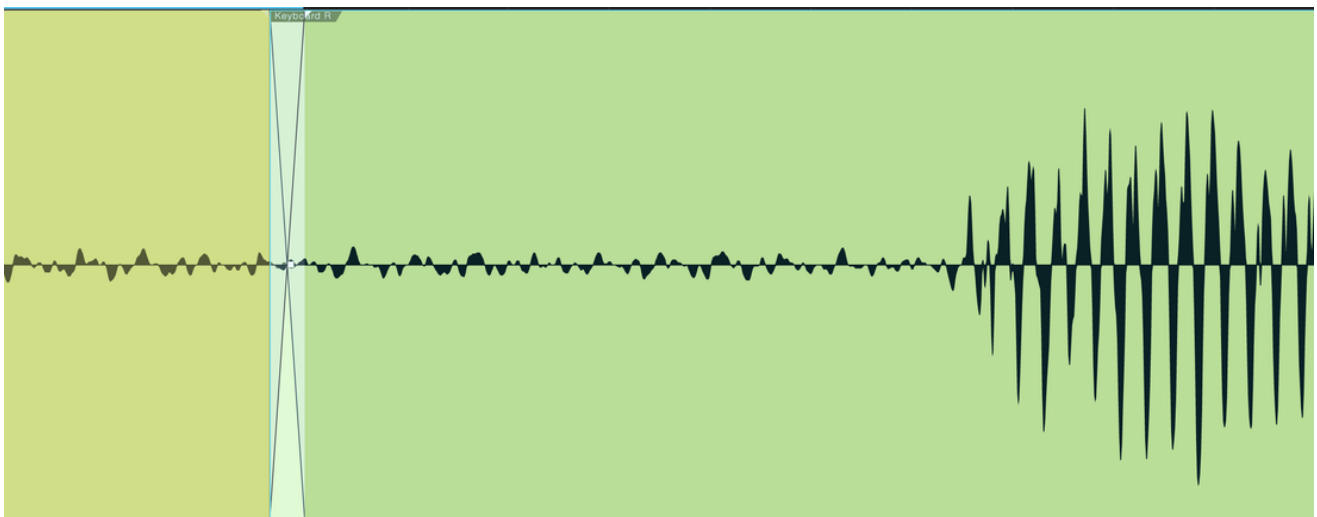
This is a good point, especially if it is a substantial break between sections. But "silence" is not without sound during rests. A piano keeps reverberating even if the pedal is up, especially if it has aliquot resonance (undamped sympathetically vibrating strings). And there's always sound from room reverberation. Therefore, it's important to match up the low-level waveforms and to crossfade the gain envelope between the two clips, as shown below. Some trial and error in splicing is always necessary here, in order to avoid hearing any small clicks or pops and any noticeable changes in background reverberation.



Splicing two clips at a silent point



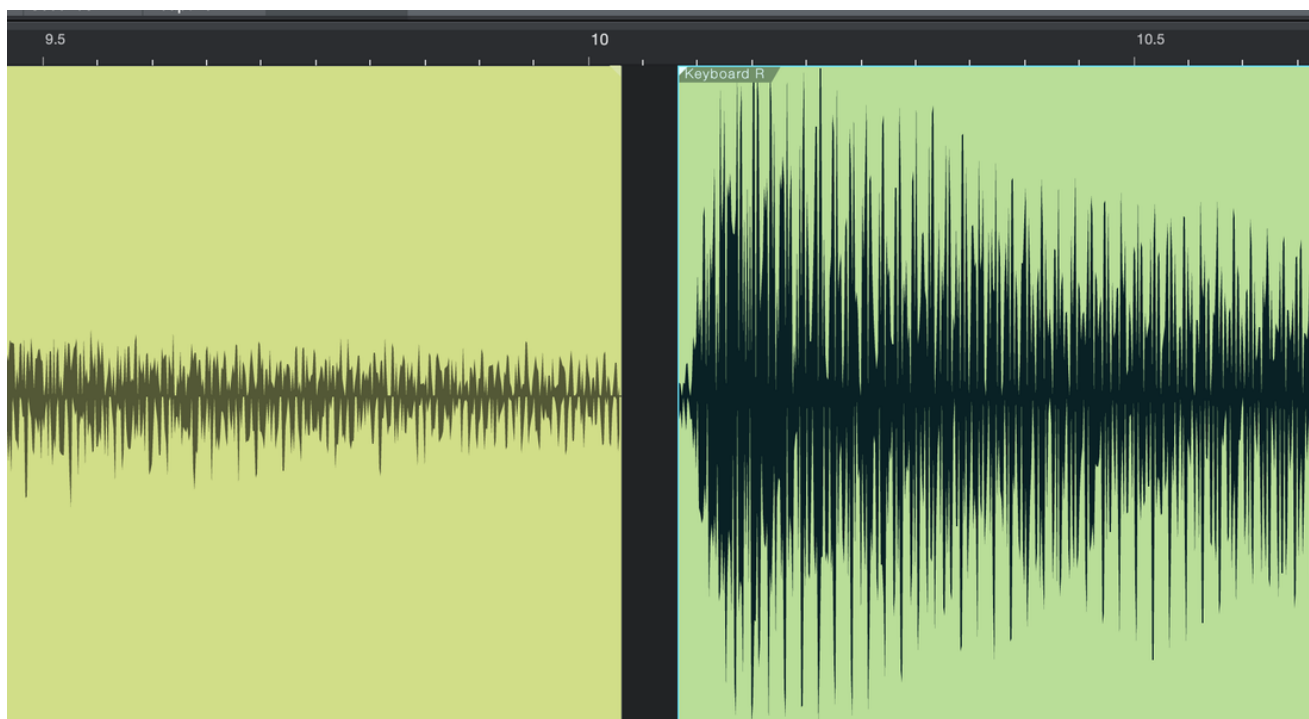
Overlapping and matching waveforms



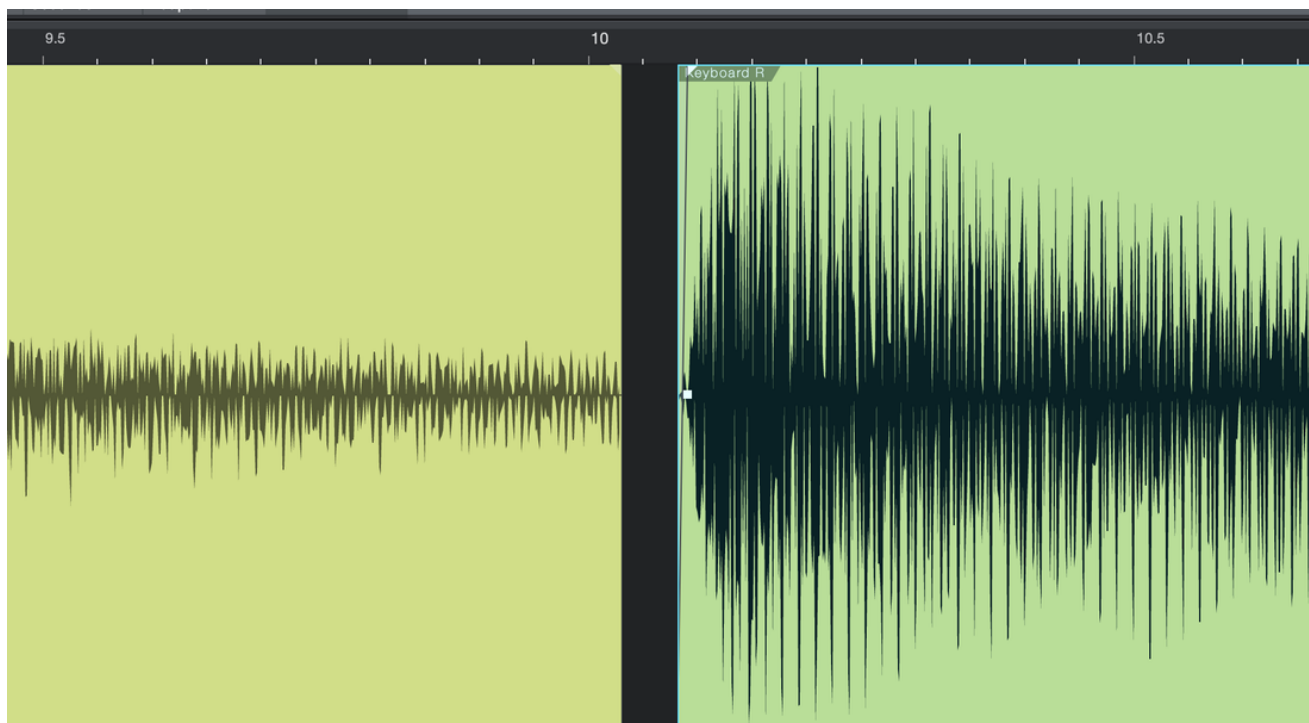
Creating a crossfade (X)

2. Splicing just before a conspicuous attack

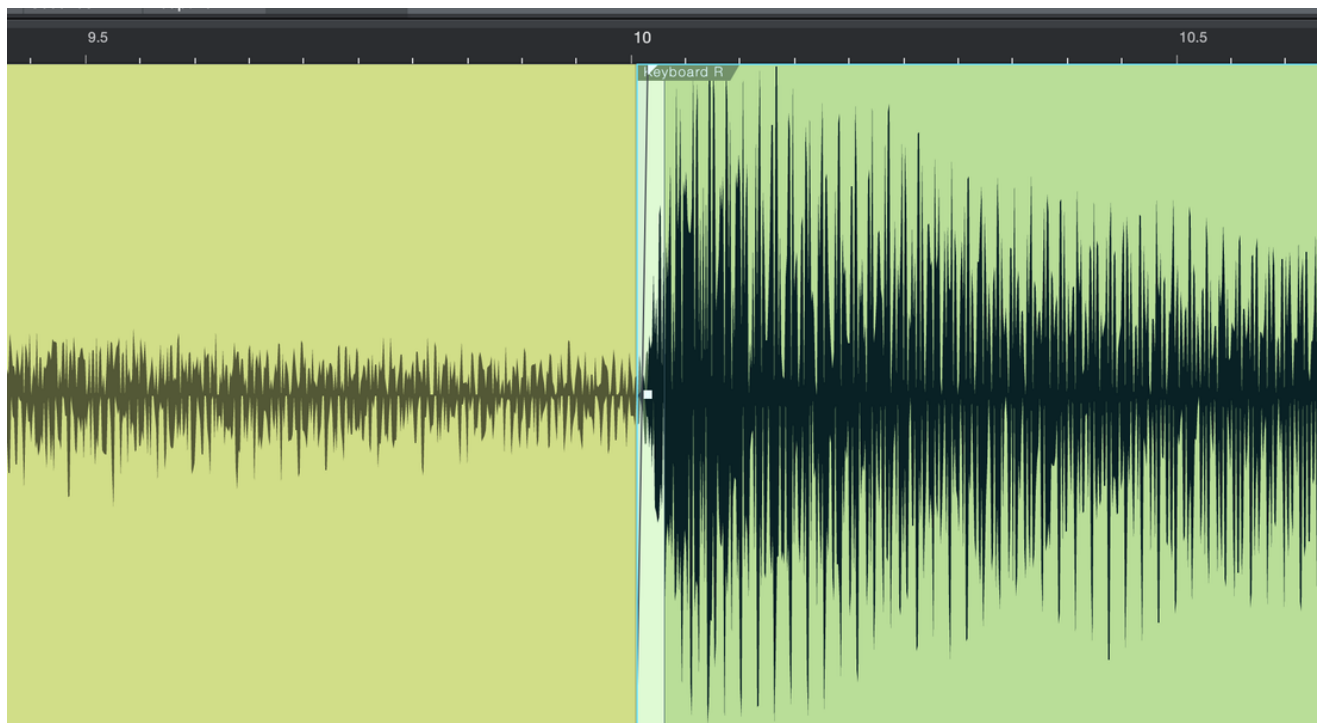
This is a point in the music when there is a large dynamic change – a much louder note suddenly occurs. Preferably, this dynamic change is accompanied by a change in tempo. (If there is no tempo change, then it is important for the performer to match the tempo as best as possible between the two musical sections.) Again, some trial and error in splicing is necessary, in order to avoid hearing any small clicks or pops and to assure that the transition sounds natural.



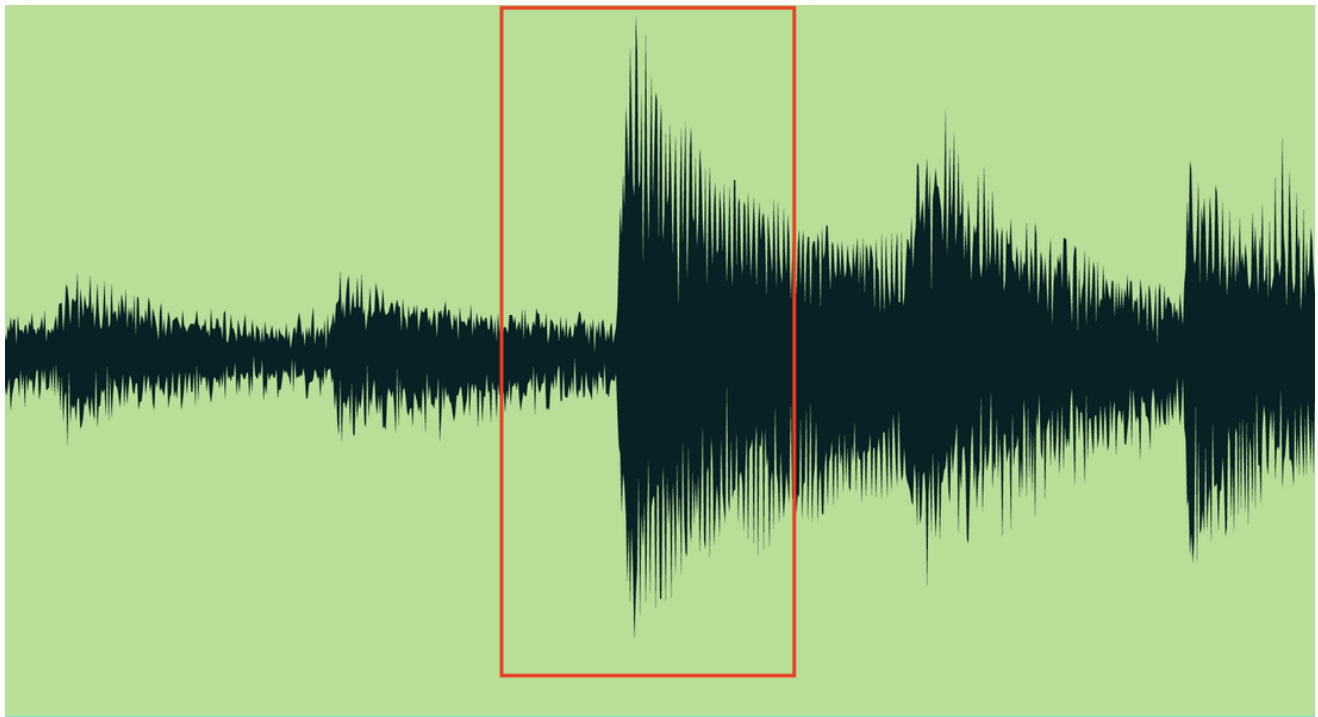
Splicing two clips at an attack point



Creating 10ms fade-in just in front of the attack note



Placing attack clip at just the right time point (overlapping clips)



"Bouncing" (merging) the two clips into one clip

Here are some parting wise words from pianist **Paul Cantrell** on recording and editing:

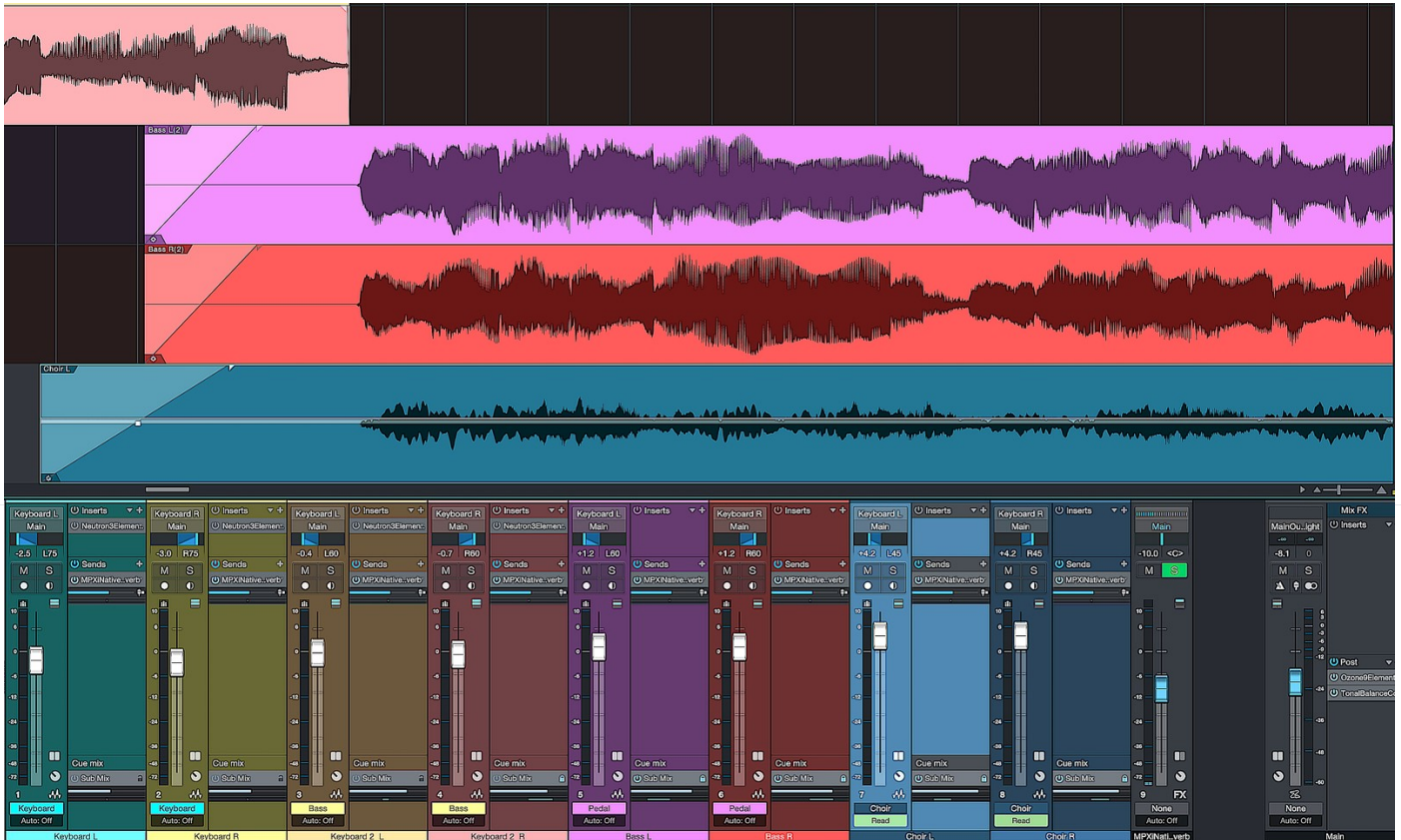
We pianists are far less consistent in tempo and dynamics than we believe — especially when starting in the middle of a passage. What we play depends on what we've been playing, and the sort of momentum we've built up. For both these reasons, when you have a mistake you want to correct or a passage you want to redo, it works *far* better to start well in advance of the passage in question. Work your way into it; never start playing on the exact note you want to splice — even if it's easier to start there, even if it's the start of a new section, and even if there's a rest before it. Back up and work your way in. You'll save yourself a bunch of heartache later!

So, how did I do in my recording and editing of Handel's Largo ? Well, to be fair, it was my first effort in my home music studio, and it was a good opportunity to experiment with and get familiar with my PreSonus Studio One digital audio workstation. The recording contains three splices – two were satisfactory, and one was not ! I'm sure you'll hear the “bad” splice if you listen closely to the recording. With a lot of practice under my belt since that recording, I have, fortunately, improved my editing skills.



Recording of Handel's Largo

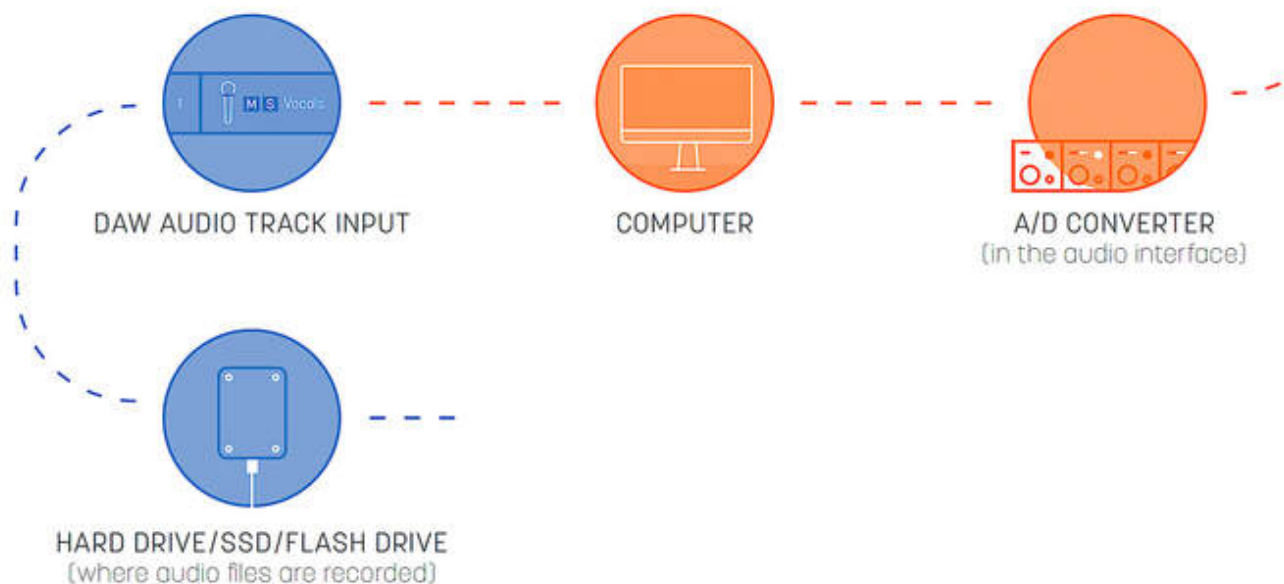
B2. DAW - Recording



Digital Audio Workstation (DAW)

In the very first chapter of this ebook, I drew up a block diagram of the **basic elements** of a home music recording studio. Without question, the heart of the system is the computer application called the **Digital Audio Workstation (DAW)**. This digital audio workstation is essentially the home studio's software version of the mixer console – that large desk of electronic hardware, sometimes called a sound board – found in professional music recording studios. So, it is no surprise that the GUI of the DAW looks just like a real mixer console. You can see the DAW interface in the screen capture above, with its channel strips, controls, meters, faders, etc. There are many well-known DAW applications, ranging from 'free' applications like Apple GarageBand that come packaged with your computer to expensive applications like Avid Pro Tools that are used in professional studios. In a previous chapter, I introduced the DAW that I use in my home studio –the **PreSonus Studio One**. For the next few posts, I will outline the basic 'work flow' of the DAW, starting in this post with the DAW input section used for the recording process.

The signal flow in the DAW is important to understand in order to utilize the software properly. Unlike hardware components where you can visualize the signal flow in a system by the cables connecting the parts, signal flow in the software programming is not readily apparent to the user. Fortunately, simple block diagrams of the signal flow in a DAW are available for download from **iZotope** . The signal flow chart for the DAW input section used for recording is shown below.



In the computer, the digital audio signal enters the DAW audio channel input section, where input gain can be trimmed and a phase shift can be applied.

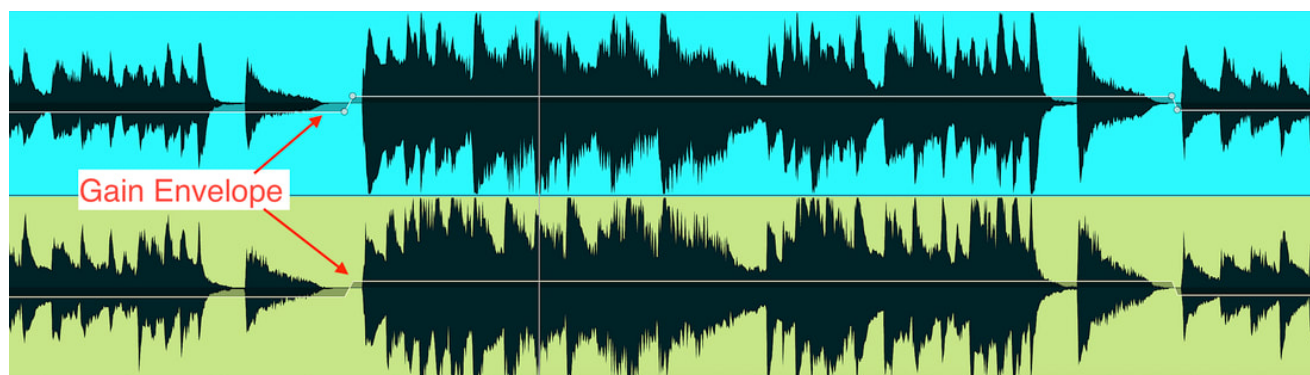


The phase Φ switch is used to shift the voltage waveform phase by 180 degrees. This has the effect of inverting the voltage waveform. This is done if two channels (say from two mics on the same sound source) have voltage waveforms that nearly cancel each other when mixed (added) together. So, engaging the 180-degree phase shift on one of these channels will mostly “fix” this waveform cancellation problem.

It should be noted that the primary adjustment for recording level is usually done back at the analog pre-amplifiers in the **Audio Interface Unit**. But the level of the digital signal can be trimmed using the input gain control at this point. Setting the appropriate input level for recording was reviewed in the prior post **Digital Signal Levels**. The audio digital waveforms are then recorded to the external solid-state drive (SSD). The SSD is the “final” destination of the signal in the recording process.

Clip Gain

An extremely useful editing feature available in most professional DAWs is the clip gain adjustment tool. A custom drawn gain envelope modifies the amplitude level of the stored digital audio waveforms during playback. This process is “non destructive”, in that the recorded (stored) digital waveforms are not altered. The gain envelope can be drawn with as much detail as desired and applied over an entire clip (event) or down to individual notes. It can also be used to surgically remove an undesired sound from a track, although this is probably better accomplished using an audio repair tool specifically meant for doing this, such as the **iZotope RX 10 plug-in**. In the following example, the gain envelope is used to increase the dynamic contrast between two musical phrases.



[Click on window to view enlarged image of Gain Envelope](#)

B3. Dynamic Microphone



I haven't used microphones in my home music studio yet, since I've focused solely on my Yamaha digital keyboard as the sound source. If I wish to record another instrument, such as voice, violin, acoustic guitar, flute, etc., I'll need to incorporate microphones in my signal chain. In this chapter, I'll give a brief overview of microphones in general, and discuss the dynamic microphone in particular.

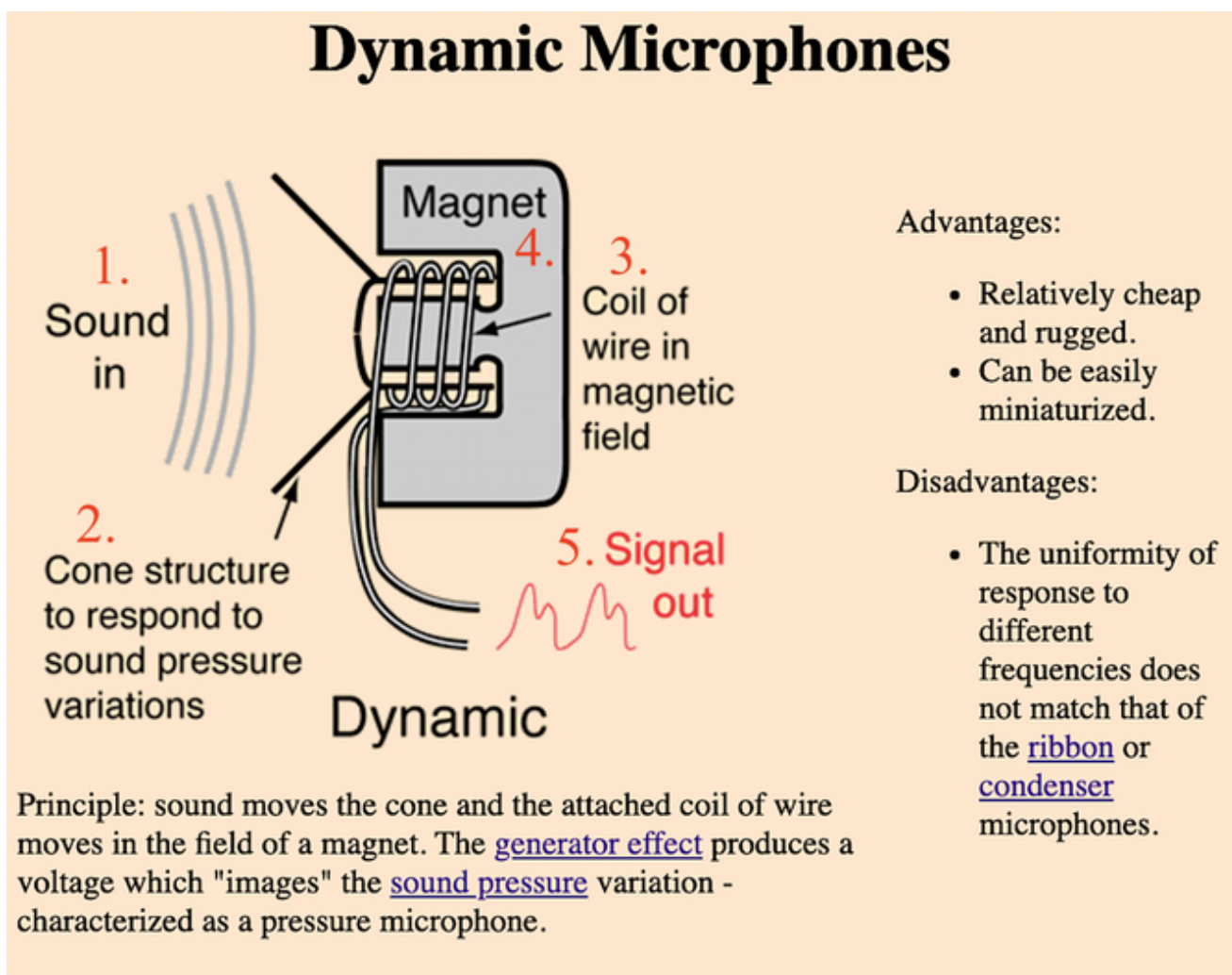
A microphone is a transducer, converting sound wave energy to electrical energy. The sound wave is a longitudinal pressure wave (compressions and rarefactions of the air molecules) that causes a thin diaphragm in the microphone to vibrate. There are basically two different mechanisms of electromagnetic physics that can convert the mechanical vibration of the diaphragm into an electrical signal:

(1) the electromotive force (EMF) (which is actually a voltage) created by the movement of electric charge in a magnetic field, and (2) the voltage across a charged two-conductor electric capacitor created by the movement of one of the conductors. Varying magnetic EMF is the operating principle of “dynamic” microphones, and varying capacitor voltage is the operating principle of “condenser” microphones. Let’s take a look at one type of dynamic microphone below, simply called a dynamic microphone. Another type of dynamic microphone, called the ribbon microphone, and the condenser microphone will be discussed in subsequent chapters.

Dynamic Microphone

The microphone (mic) shown in the photo at the top of this post is a **Shure SM58 dynamic mic**. The Shure SM58 has been one of the most popular microphones in the music industry for decades, particularly for live vocal performances.

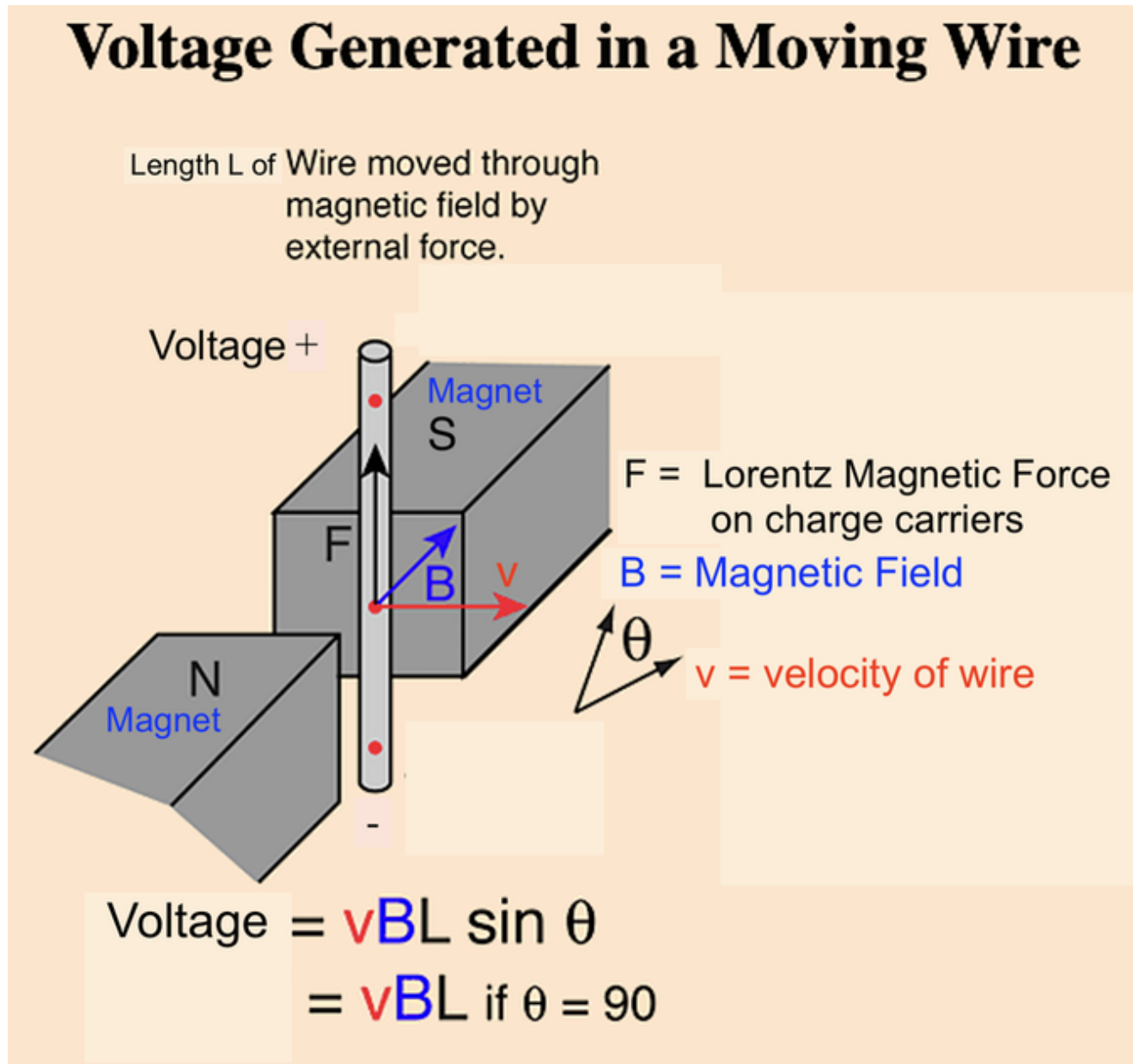
The “inner workings” of a dynamic mic are shown in the figure below.



adapted from C.R. Nave <http://hyperphysics.phy-astr.gsu.edu/>

The sound pressure waves (1.) strike the diaphragm cone (2.) and cause it to vibrate. The cone is attached to a metal wire coil (3.) that can move over a cylindrical magnet pole (4.). As the coil moves, the mobile charge carriers in the wire feel a force, the Lorentz magnetic force, creating an EMF voltage across the terminals of the coil. This voltage is the source of the electrical signal in the microphone output circuit (5.) .

Let's look more closely at the EMF voltage generated by this moving wire coil. The Lorentz magnetic force that acts on the mobile charge carriers in a wire is shown in the figure below.



adapted from C.R. Nave <http://hyperphysics.phy-astr.gsu.edu/>

Here's my simple "derivation" of the EMF voltage generated by the moving wire coil in the dynamic microphone:

This vector magnetic force per unit charge is given by:

$$\vec{f}_{mag} = \vec{v} \times \vec{B} \quad (1)$$

In the cylindrical coordinates of the coil and magnet system shown in the dynamic mic diagram above,

The velocity vector of the coil is along the z-axis:

$$\vec{v} = v \hat{z} \quad (2)$$

The magnetic field vector between the north and south poles of the magnet is in the radial direction and is constant everywhere along the wire:

$$\vec{B} = B \hat{r} \quad (3)$$

So, by the right-hand-rule of the cross product of vectors in (1), the magnetic force vector is constant and circumferential (tangent) everywhere along the wire:

$$\vec{f}_{mag} = v B \hat{\phi} \quad (4)$$

The EMF voltage generated at the output terminals of the coil is given by:

$$V = \oint \vec{f}_{mag} \cdot d\vec{l} = \oint v B dl = 2 \pi R N v B \quad (5)$$

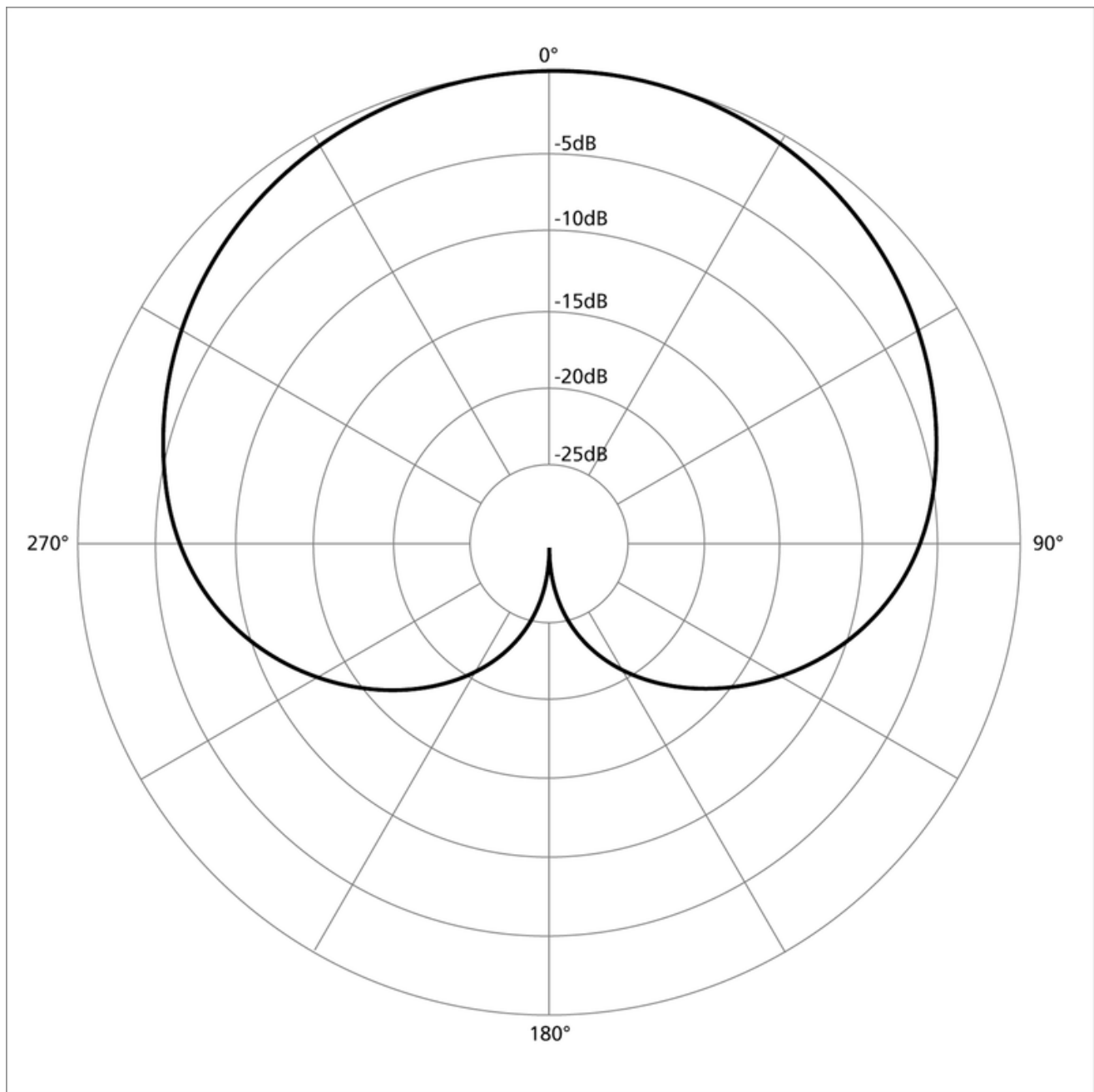
where R is the radius of the coil and N is the number of turns of the coil.

The pressure of the sound wave causes the cone to vibrate about its rest position. Consequently, the coil of wire moves with positive and negative velocities, and the generated voltage in (5) is proportional to the velocity. This voltage waveform therefore “images” the pressure variations of the input sound wave and constitutes the output signal of the microphone.

The important characteristics of a microphone include:

1. Polar response pattern
2. Frequency response
3. Output signal level
4. Output resistance (impedance)

Dynamic mics usually have a “cardioid” polar response pattern, as shown here:

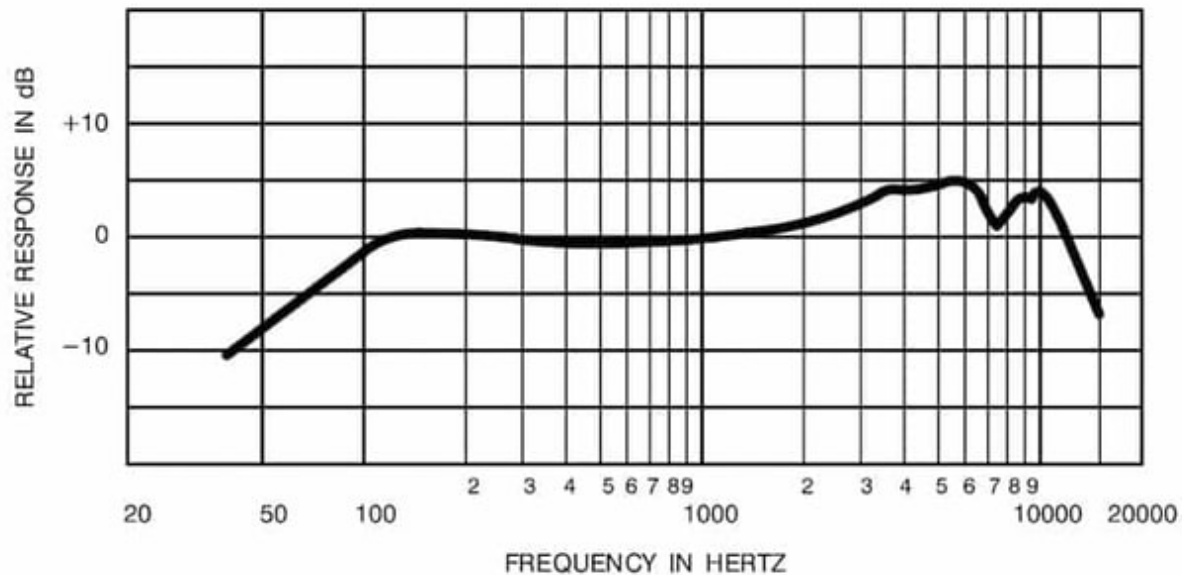


Cardioid Pattern by Galak76

With a cardioid pattern, the mic picks up sound well in a forward-looking direction while rejecting sound from its rear. This pattern is good for isolating sound sources on a recording.

Because of the fairly heavy diaphragm and coil, movement of the coil is ‘sluggish’ and restricted,

which in turns decreases its high-frequency and transient response. A typical frequency response curve for a dynamic mic is given here:



Shure SM58 Microphone Frequency Response

Typical microphone output levels vary in the range 1 – 10 mV and their output impedances range from 50 Ohm to 600 Ohm . For the **Shure SM58**, the spec sheet gives an output level of 1.6 mV and an output impedance of 300 Ohm.

Dynamic mics have a number of key advantages that include,

- Rugged and inexpensive
- Don't require external power (phantom power)
- Moderate frequency response

The sound produced from a dynamic mic is characterized as mellow and well rounded.

There is another kind of “dynamic” microphone, called a ribbon mic, that operates by sound waves vibrating a thin, delicate metallic ribbon in a magnetic field. We'll take a closer look at the ribbon microphone in the next chapter.

B4. Ribbon Microphone

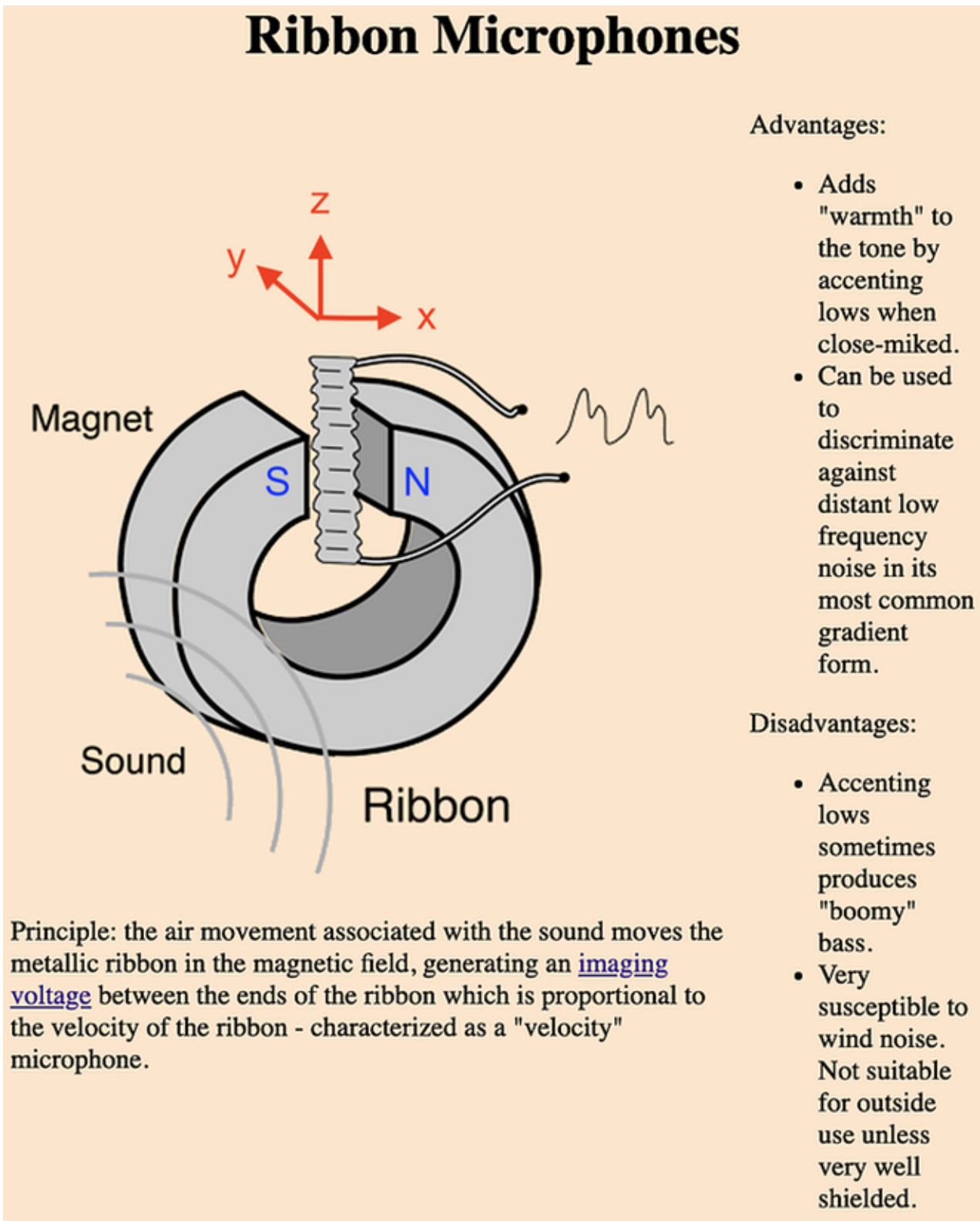


Photo: Jim Merithew/Wired.com

A microphone is a transducer, converting sound wave energy to electrical energy. The sound wave is a longitudinal pressure wave (compressions and rarefactions of the air molecules) that causes a thin diaphragm in the microphone to vibrate. As presented in the previous chapter on **Dynamic Mics**, this vibrating diaphragm is attached to a metal wire coil that moves in the presence of a magnetic field, generating an electromotive force (EMF) voltage. There is a related type of “dynamic” microphone, called the Ribbon Microphone, where a very thin, metallic diaphragm moves back and forth within a permanent magnetic field, also generating an EMF voltage.

The microphone (mic) shown in the photo above is an **Audio Engineering Associates (AEA) R84 ribbon mic**. The AEA R84 ribbon mic is a popular choice of many studio engineers for its warm and rich sound, reminiscent of the classic RCA ribbon mics of the 1930s.

The "inner workings" of a ribbon mic are shown in the figure below.



Sound pressure waves strike a thin, corrugated aluminum ribbon, usually from both the front and back sides. The thickness of this ribbon is only 2 microns (micrometers), and its dimensions are on the order of 5 mm width x 60 mm length. The difference in pressure between front and back (pressure gradient) causes the ribbon to move inward and outward from its rest position. The very thin, lightweight ribbon can react swiftly and accurately to the incoming sound wave pressure differential. The ribbon moves with velocity at right angles to the magnetic field lines between the magnet poles. The mobile charge carriers in the metallic sheet feel a force, the Lorentz magnetic force, creating an EMF voltage across the terminals of the ribbon. This voltage is the source of the electrical signal in the microphone output circuit.

Let's look more closely at the EMF voltage generated by this moving metallic ribbon. The Lorentz magnetic force acts on the mobile charge carriers in the metallic ribbon.

This vector magnetic force per unit charge is given by:

$$\vec{f}_{mag} = \vec{v} \times \vec{B} \quad (1)$$

In the rectangular coordinates of the ribbon and magnet system shown in the ribbon mic diagram above,

the velocity vector of the ribbon is along the y-axis:

$$\vec{v} = v \hat{y} \quad (2)$$

The magnetic field vector between the north and south poles of the magnet is in the negative x-axis direction and is considered to be constant everywhere along the ribbon:

$$\vec{B} = -B \hat{x} \quad (3)$$

So, by the right-hand-rule for the cross product of vectors in (1), the magnetic force vector is constant and directed along the z-axis :

$$\vec{f}_{mag} = v B \hat{z} \quad (4)$$

The EMF voltage generated at the output terminals of the ribbon is given by:

$$V = \oint \vec{f}_{mag} \cdot d\vec{l} = \oint v B dz = v B L \quad (5)$$

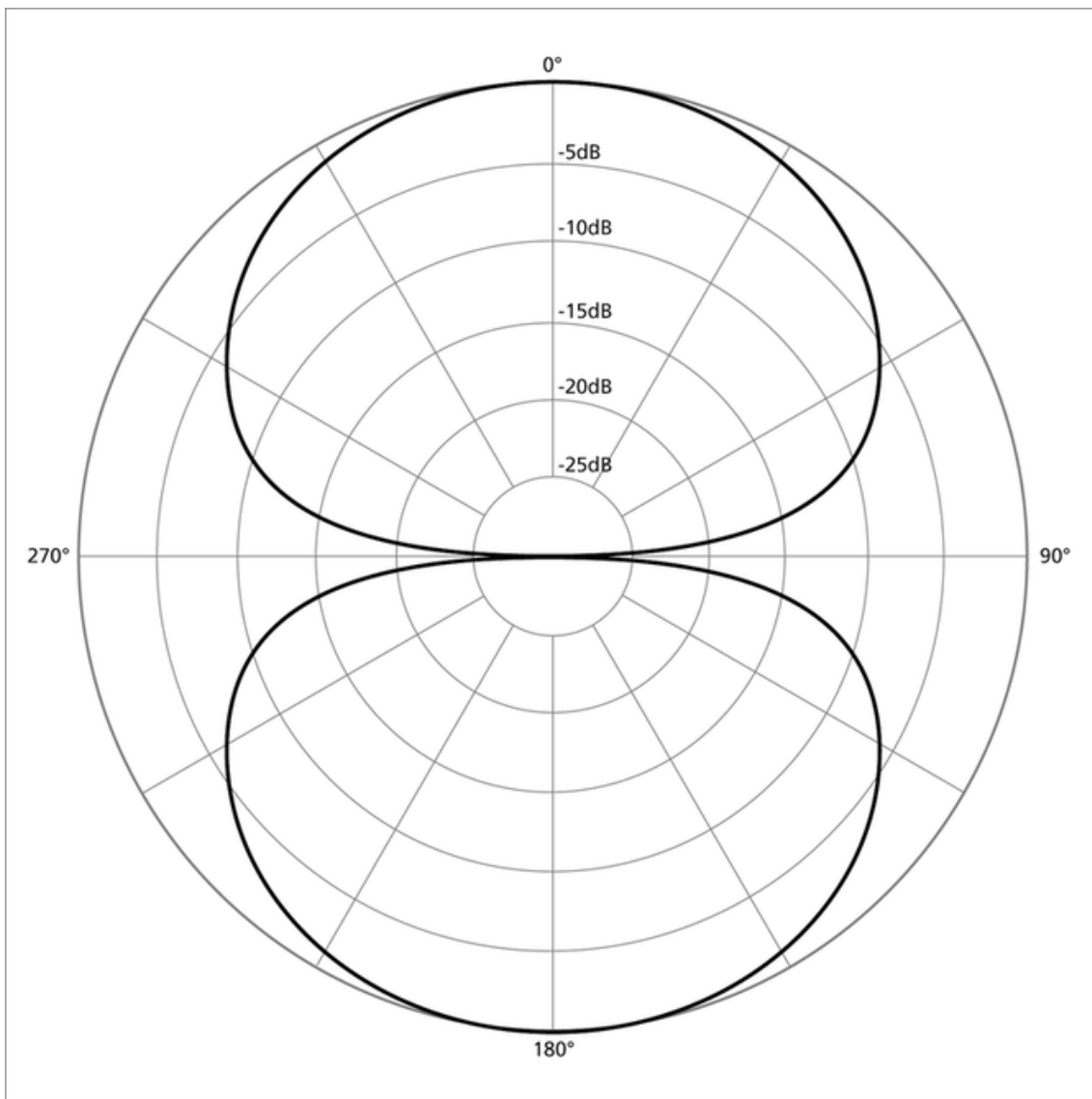
where L is the length of the ribbon.

The differential pressure of the sound wave causes the ribbon to vibrate about its rest position. Consequently, the ribbon moves with positive and negative velocities, and the generated voltage in (5) is proportional to the velocity. This voltage waveform therefore “images” the pressure variations of the input sound wave and constitutes the output signal of the microphone.

The important characteristics of a microphone include:

- 1.Polar response pattern
- 2.Frequency response
- 3.Output signal level
- 4.Output resistance (impedance)

Ribbon mics usually have a “bi-directional” polar response pattern, as shown here:



Bi-Directional Pattern by Galak76

Ribbon microphones, by the nature of their design, are pressure-gradient microphones. Both the front and rear of the ribbon are exposed equally to sound pressure variations. With sound waves approaching from the front of the mic, there is a phase shift and amplitude change between the pressure waves that are incident on the front and back of the ribbon itself. The difference in sound pressure between the two sides causes the ribbon to move, and a mic signal is generated. The same thing happens when sound waves approach from the rear of the mic. But when sound waves approach the mic from the sides, there will be equal pressures on the front and back of the ribbon. Therefore, there is no motion of the ribbon and no output signal generated. With this bi-directional ("figure-8") pattern, the mic picks up sound well in both forward- and reverse-looking directions while rejecting sound from its sides. This pattern can be used to good effect, such as in making a stereo microphone pair (the Blumlein technique) - more on this in an upcoming chapter.

The directionality property of this microphone leads to a very interesting frequency response characteristic, called the proximity effect . This effect is discussed next.

Proximity Effect

The proximity effect is the increase in low-end frequency response in a pressure gradient microphone as the microphone is brought closer to the sound source.

The pressure difference between the front and back of the ribbon causes it to move. This difference is created mainly by the phase and amplitude changes of the sound wave as it propagates from the front of the mic to the back (and vice versa).

The sound-wave phase shift is frequency dependent, with the pressure gradient sensitivity increasing at a rate of approximately 6 dB per octave. To compensate for this increasing sensitivity, the ribbon mounting structure is damped to "flatten" the microphone frequency response, *i.e.*, the damping causes a sensitivity decrease of roughly 6 dB per octave.

On the other hand, sound-wave amplitude change is NOT frequency dependent. The amplitude decrease of an outwardly propagating sound wave generated by a point source is given by an inverse square law with distance traveled. As the distance between the sound source and the microphone gets smaller, the amplitude difference of the pressure wave between the front and back of the ribbon becomes larger. Below, I use some simple algebra to demonstrate this effect .

This amplitude difference effect can be demonstrated using some simple math involving the inverse square law for pressure wave amplitude P versus distance from the source r ,

$$P = \frac{A}{r^2}$$

A is a constant. We can write the pressure amplitude P at the front and back of the ribbon as,

$$P_{front} = \frac{A}{r_o^2}$$

$$P_{back} = \frac{A}{(r_o + \Delta r)^2}$$

where r_o is the distance from sound source to the front of the ribbon, and Δr is the additional distance that the wave must travel from the front of the ribbon to the back of the ribbon. The change in amplitude, normalized to the amplitude of the wave incident on the front side of the ribbon, is given by,

$$\frac{\Delta P}{P} = \frac{P_{front} - P_{back}}{P_{front}} = 1 - \frac{1}{\left(1 + \frac{\Delta r}{r_o}\right)^2}$$

Let's look at this sound pressure level (SPL) difference on a decibel scale,

$$P_{back} = \left[1 - \left(1 - \frac{1}{\left(1 + \frac{\Delta r}{r_o}\right)^2} \right) \right] P_{front} = \frac{1}{\left(1 + \frac{\Delta r}{r_o}\right)^2} P_{front}$$

$$P_{back} (dB SPL) = -20 \log \left[\left(1 + \frac{\Delta r}{r_o} \right)^2 \right] (dB) + P_{front} (dB SPL)$$

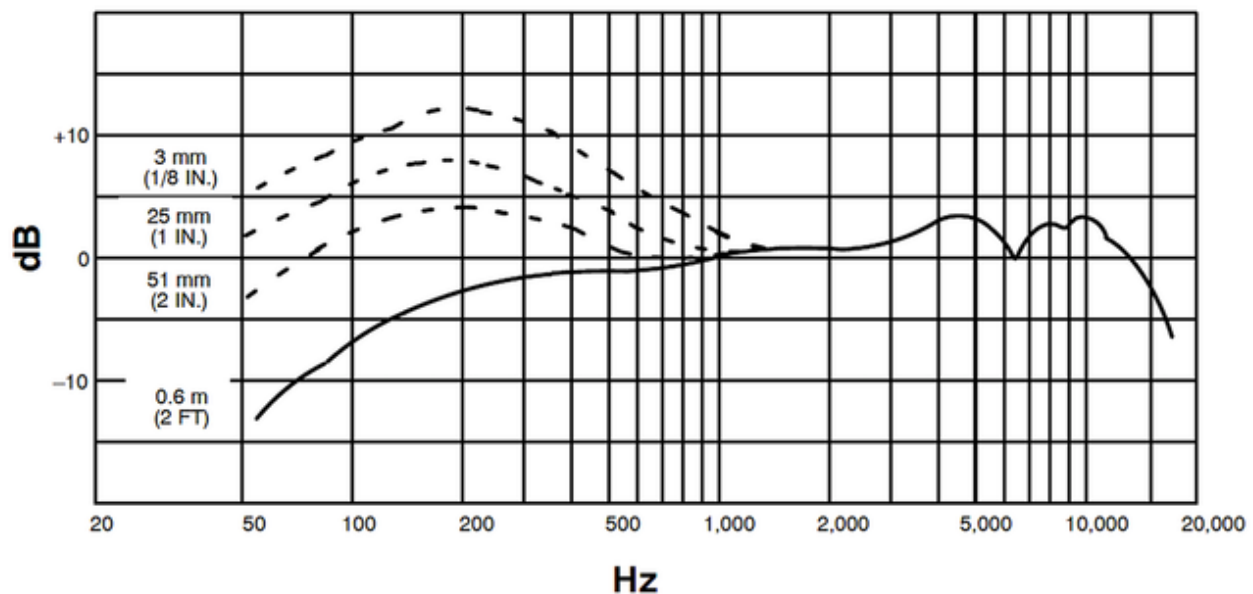
$$P_{front} (dB SPL) - P_{back} (dB SPL) = 20 \log \left[\left(1 + \frac{\Delta r}{r_o} \right)^2 \right] (dB)$$

$$= \begin{cases} \sim 0 \text{ dB} , & r_o \gg \Delta r \\ \sim 12 \text{ dB} , & r_o \rightarrow \Delta r \end{cases}$$

So, when the microphone is far from the sound source, $r_o \gg \Delta r$, there is negligible change in the sound pressure level between the front and back of the ribbon. As the microphone is brought close to the sound source, in the limit $r_o \rightarrow \Delta r$, there is significant change in the sound pressure level between the front and back of the ribbon – on the order of 12 dB difference !

With the ribbon mic positioned close to the sound source, the pressure difference across the ribbon at the low and low-mid frequencies is dominated by wave amplitude changes. This amplitude “boost” in the bass frequencies is called the **proximity effect**. At mid and higher frequencies, the wave phase-change effect on pressure gradient is rapidly increasing with frequency (approx. 6 dB/octave) and takes over as the dominant factor in pressure gradient. (Recall that in order to keep the frequency response somewhat “flat” as frequency increases, there is damping of the response by the ribbon mount structure at a rate of roughly -6 dB/octave.)

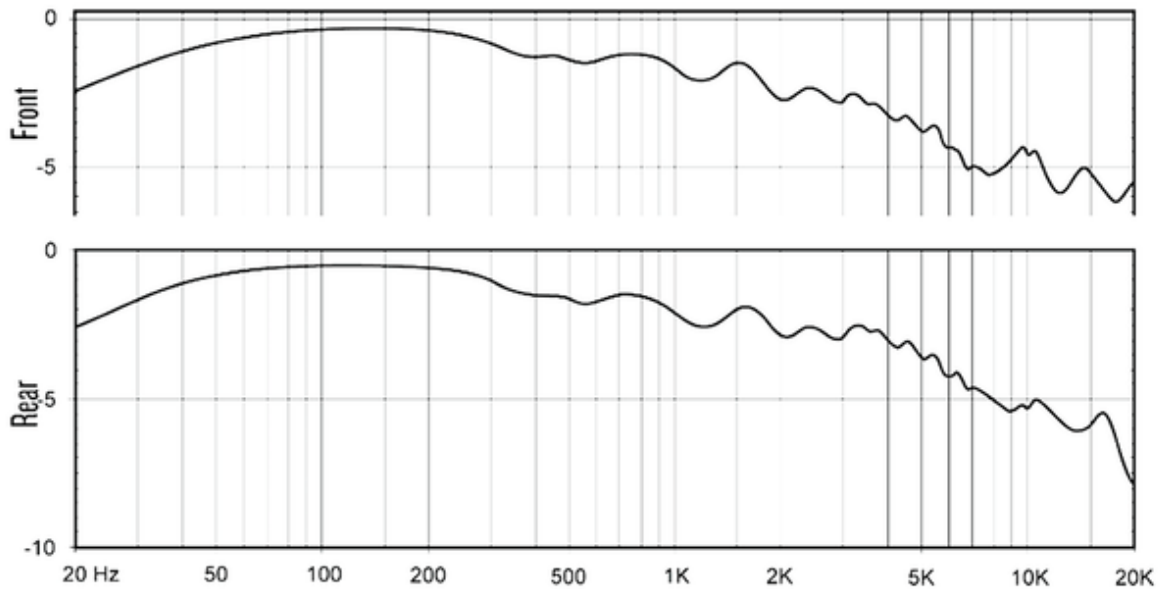
The frequency response curve for the **Shure BETA 57A pressure-gradient microphone** shows very nicely the “bass boost” from the proximity effect. Microphone placements range from a large distance from source (2 ft.) to a very short distance from source (1/8 inch).



Proximity Effect in Shure Beta 57A Microphone

The frequency response curve for the **AEA R84 ribbon microphone** is shown below. It is not stipulated in the specifications sheet what the distance is between sound source and the ribbon microphone in this frequency response graph.

FREQUENCY RESPONSE



AEA R84 Ribbon Microphone (0 dB = -55 dBV output at 1 Pascal (94 dB SPL))

The proximity effect with its low-end boost gives this ribbon microphone its desired rich, warm, even dark characteristic that is well suited for voice and solo instruments. By close mic'ing the instrument, the low-end can be effectively added to a thin mix. As usual, caution must be exercised to avoid a muddy low-end in the mix, so use proper mic positioning for recording. You don't want to have to "fix in the mix" with EQ, a practice very much frowned upon !

Finally, here are some notes on the output signal level and impedance of a ribbon mic. Because the EMF voltage generated by the ribbon is much less than that generated by the coil in a dynamic mic, a "step-up" transformer is typically used on the output. This transformer boosts the output voltage level, but unfortunately also increases the output resistance. (The transformer also prevents phantom power (48 VDC) from being accidentally applied across the ribbon, which could destroy the ribbon.) The spec sheet for the **AEA R84 ribbon mic** lists an output voltage (sensitivity) of 2.5 mV per Pascal of pressure into an unloaded circuit, and an output impedance of 270 Ohms. Due to the low output voltage level and relatively high, frequency-dependent output impedance, ribbon mics are often used with their own specially designed pre-amps. A high-input-impedance, high-gain ribbon-specific preamp will really bring out the charm and character of a passive ribbon microphone.

In the next chapter, we'll take a look at the studio workhorse microphone – the condenser mic.

B5. Condenser Microphone



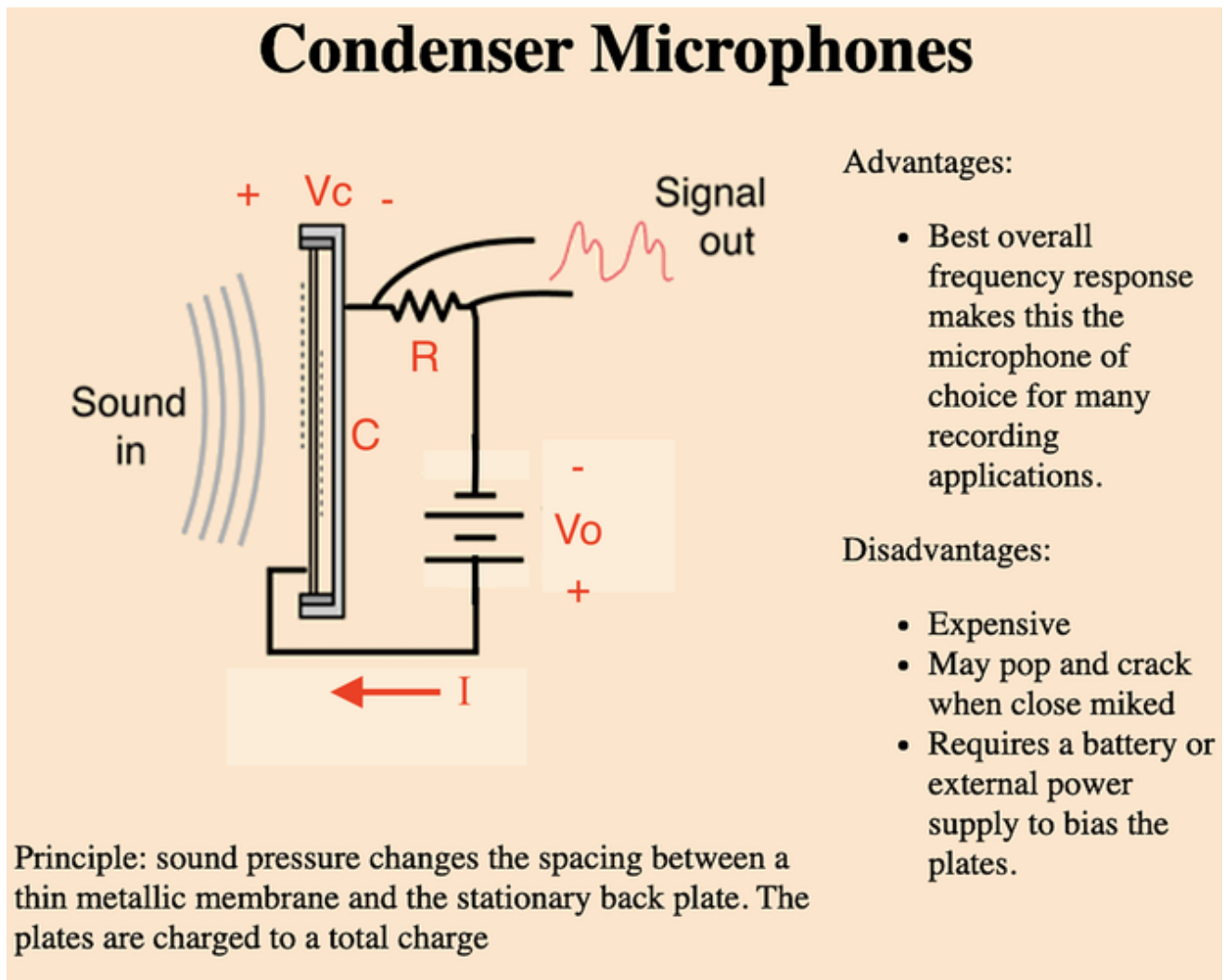
Neumann TLM 103 Condenser Microphone © Neumann.Berlin

A microphone is a transducer, converting sound wave energy to electrical energy. The sound wave is a longitudinal pressure wave (compressions and rarefactions of the air molecules) that causes a thin diaphragm in the microphone to vibrate. As we've seen in the two previous chapters on **dynamic microphones** and **ribbon microphones**, a metal wire coil or metallic ribbon moves in the presence of a magnetic field, generating an electromotive force (EMF) voltage. There is a third type of microphone, the condenser microphone, that creates an output signal voltage by another physical process – the vibration of a metallic diaphragm that is part of an electrostatic capacitor in an electric circuit.

Condenser microphones are very popular in the music studio owing to their clear, bright sound over an extended frequency range. They also have high sensitivity and accurate transient response. Condenser mics contain “active” circuits, requiring a voltage source to bias the electric capacitor and output amplifier circuit. Supplying this DC bias voltage through the microphone XLR cable is usually done by applying “phantom power” from the mixer console or audio interface.

The microphone shown in the photo above is a **Neumann TLM103 large-diaphragm condenser mic** – a mainstay of most studios around the world. The Neumann TLM103 mic possesses a clear voicing with a wide presence boost for frequencies above 5 kHz. This mic is very well suited to bring out vocals and solo instruments in the mix.

The “inner workings” of a condenser mic are shown in the figure below.



adapted from C.R. Nave <http://hyperphysics.phy-astr.gsu.edu/>

Sound pressure waves strike a thin, metallic diaphragm that comprises the front conductor of a two-conductor electric capacitor. The difference in pressure between front and back (pressure gradient) causes the metallic diaphragm to move inward and outward from its rest position. The very thin, lightweight diaphragm can react swiftly and accurately to the incoming sound-wave pressure differential. As the front conductor moves, the voltage across the capacitor changes, creating a time-varying output signal voltage that “images” the pressure variations of the input sound wave.

Let’s look more closely at the output voltage generated by the varying capacitance of the biased capacitor in the circuit diagram above.

The two-conductor capacitor is initially fully charged to $+Q_0$ charge on the front conductor and $-Q_0$ charge on the back conductor, so the voltage across the capacitor is given by,

$$V_C(t = 0^-) = V_0 = \frac{Q_0}{C_0} \quad (1)$$

The capacitance of a parallel-plate capacitor is given by,

$$C_0 = \frac{\epsilon A}{d} \quad (2)$$

where d is the rest distance between the movable front conductor diaphragm and the fixed rear conductor backplate. (ϵ is the permittivity of the air between the plates, and A is the surface area of one plate.) V_0 is the DC bias voltage (typically 48V from phantom power).

At $t = 0^+$, the movable front conductor diaphragm is instantaneously displaced a small amount Δx so that the distance between the conductors increases to $d + \Delta x$. The work needed to move the diaphragm is done by the force of the sound pressure wave. The energy put into the system is reflected by the increased voltage across the capacitor,

$$V_C(t = 0^+) = \frac{Q_0}{C} = Q_0 \cdot \frac{(d + \Delta x)}{\epsilon A} = \frac{Q_0}{C_0} \left(1 + \frac{\Delta x}{d}\right) \quad (3)$$

The biased resistor-capacitor circuit is now out of equilibrium, and electric currents will begin to flow that will act to discharge the capacitor. Kirchhoff’s voltage law gives the dynamic equation governing the flow of charge from the capacitor,

$$V_C - V_0 + IR = 0 \quad (4)$$

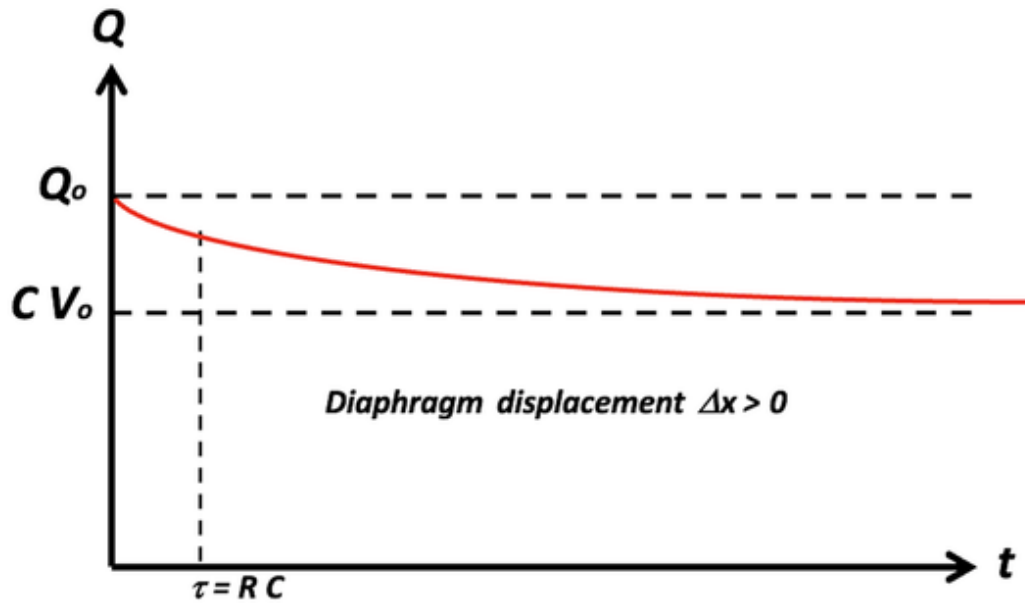
Since the current $I = \frac{dQ}{dt}$, we can arrange the equation in (4) as follows,

$$\frac{dQ}{dt} + \frac{Q}{RC} - \frac{V_0}{R} = 0 \quad (5)$$

The solution to this differential equation for the charge $Q(t)$ on the capacitor, subject to the initial condition $Q(t = 0) = Q_0$, is,

$$Q(t) = CV_0 + (Q_0 - CV_0)e^{-t/RC} \quad (6)$$

A plot of this capacitor charge Q versus time t is shown below, for a diaphragm displacement $\Delta x > 0$,



If $\Delta x = 0$, $C = C_0$ and $Q(t) = Q_0$

If $\Delta x > 0$, $C < C_0$ and $Q(t \rightarrow \infty) = C V_0 < Q_0$ the capacitor discharges

If $\Delta x < 0$, $C > C_0$ and $Q(t \rightarrow \infty) = C V_0 > Q_0$ the capacitor charges

As seen in the graph above for the capacitor discharging, the time constant, $\tau = R C$, characterizes how quickly the charge Q leaves the capacitor plate. For very large values of resistance R , the time constant τ is quite large, indicating that discharge of the capacitor is a very slow process.

Now consider the front conductor diaphragm vibrating about its rest position at audio frequencies. The diaphragm displacements Δx are caused by the air-pressure compressions and rarefactions of the incident sound wave,

$$\Delta x(t) \propto P \sin(2\pi f_s t) \quad \text{where } f_s \text{ is the frequency of the sound wave}$$

$$\text{For frequencies } f_s \gg f_{cap} = \frac{1}{2\pi\tau} = \frac{1}{2\pi RC},$$

the capacitor cannot charge or discharge quickly enough to “keep up” with the diaphragm displacements. Consequently, so long as the time constant $\tau = RC$ is very large, the charge on the capacitor plate remains near its equilibrium value at all times,

$$Q(t) \cong Q_0 \quad (7)$$

And the voltage across the capacitor is approximately,

$$V_C(t) \cong \frac{Q_0}{C(t)} = \frac{Q_0}{C_0} \left(1 + \frac{\Delta x(t)}{d}\right) = V_0 \left(1 + \frac{\Delta x(t)}{d}\right) \quad (8)$$

Finally, the signal voltage is taken as the voltage across the resistor R ,

$$V_R(t) = V_C(t) - V_0 \cong V_0 \left(1 + \frac{\Delta x(t)}{d}\right) - V_0 = V_0 \frac{\Delta x(t)}{d}$$

$$V_R(t) \cong V_0 \frac{\Delta x(t)}{d} \quad (9)$$

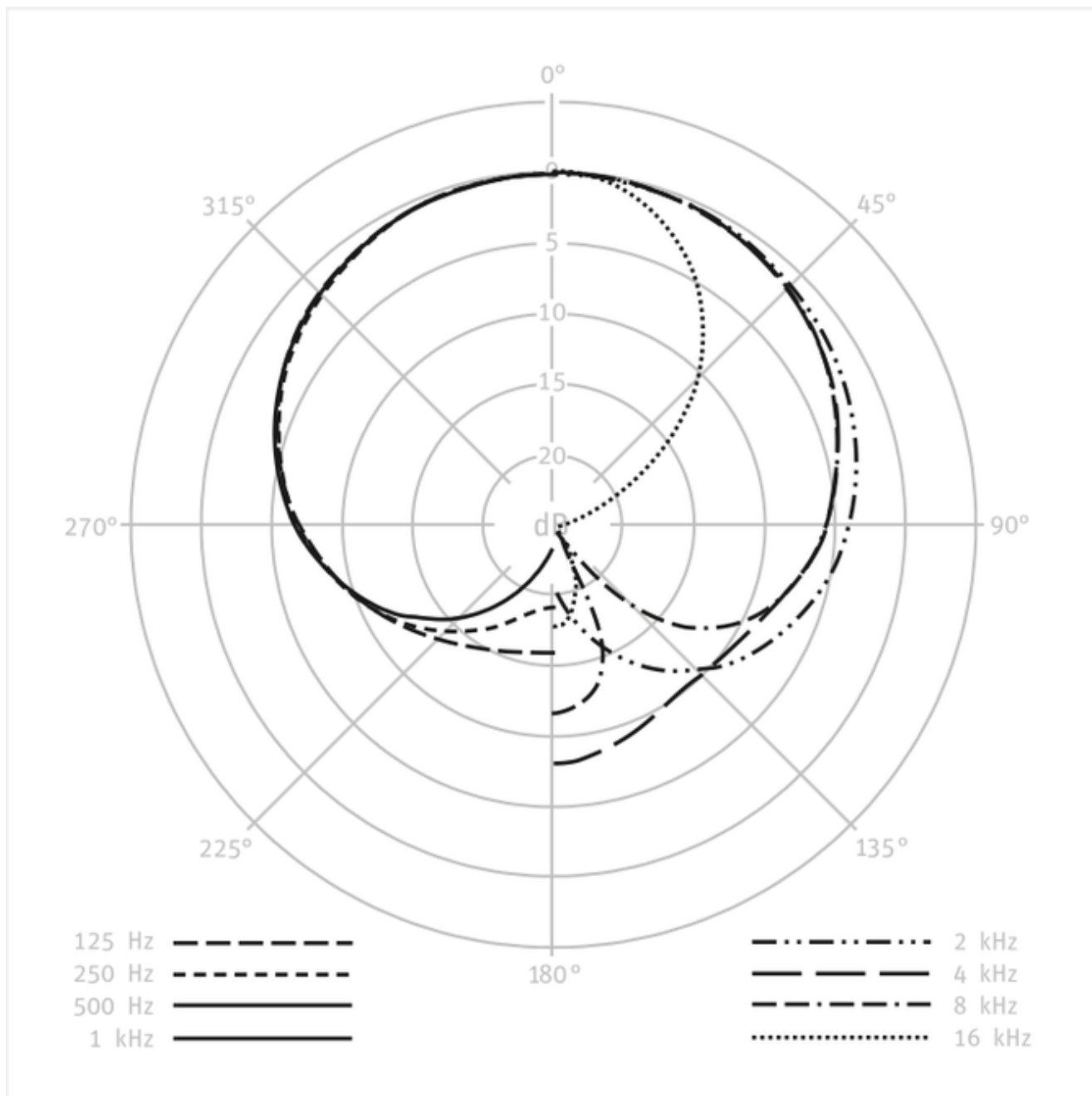
Therefore, the output signal voltage in (9) is proportional to the diaphragm displacement from its rest position.

The time-varying output voltage is taken from across the very large resistor R in the circuit. Consequently, there is a very high output impedance for the signal source. This is not a desirable situation for connecting to the pre-amp of a mixer console or audio interface, as we saw in an early chapter discussing the **voltage divider effect in audio cable circuits**. Therefore, we need to place an impedance converter (buffer amp) immediately after the RC circuit to drop the impedance of the signal source. Most modern condenser mics use a field-effect transistor (FET) amplifier to convert a high-impedance source to a low-impedance one. And there is usually some voltage gain of the signal provided by the amplifier, increasing the output signal level. The field-effect transistor circuit requires voltage biasing (48 VDC) which is usually supplied by phantom power coming from the XLR cable attached to the mic.

The important characteristics of a microphone include:

1. Polar response pattern
2. Frequency response
3. Output signal level
4. Output resistance (impedance)

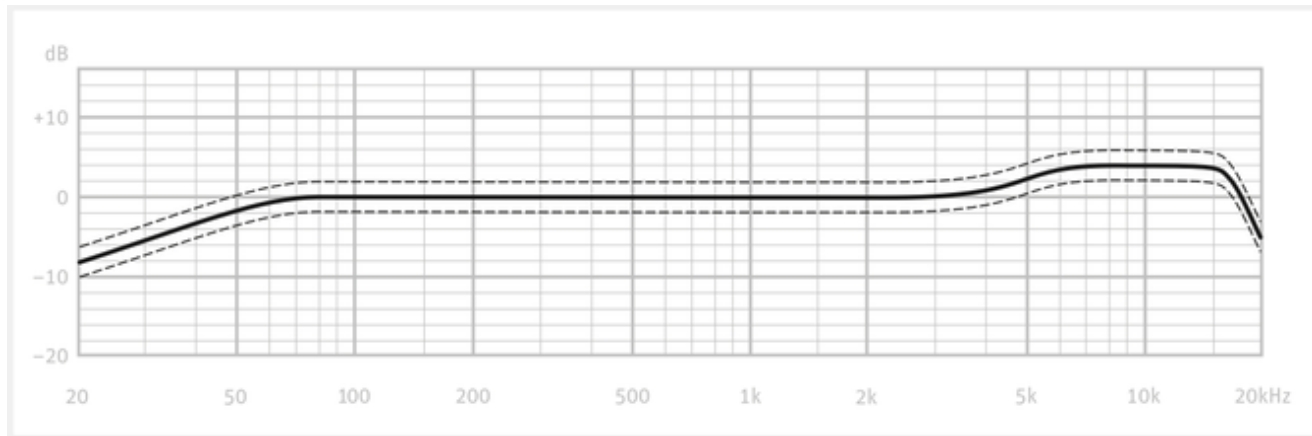
Condenser mics typically have a cardioid polar pattern, as shown here for the Neumann TLM103 mic:



Neumann TLM 103 Condenser Mic Polar Response Pattern © Neumann.Berlin

Some condenser mics have dual diaphragms, *i.e.*, two back-to-back capacitors that share a common fixed inner plate. These dual-diaphragm condenser mics exhibit multiple polar response patterns by electrically combining the signals from the two capacitors in different ways. The ability to switch between a cardioid, bi-directional, or omni-directional polar pattern makes these mics very versatile in the recording studio.

The frequency response curve for the Neumann TLM103 condenser microphone is shown below.



Neumann TLM 103 Condenser Mic Frequency Response © Neumann.Berlin

The wide presence boost from 5 kHz to 15 kHz gives this microphone its bright, clear character that is well suited for bringing out vocals and solo instruments in the mix. The very flat response from 5 kHz down to 60 Hz ensures a very immediate, uncolored sound that is true to the original. The bass roll-off below 60 Hz reduces unwanted low-end rumble and noise. In addition to using this mic for vocals and solo instruments, it is commonly employed as an ambient mic in stereo recordings of classical orchestral music. The topic of stereo microphone techniques will be the subject of the next post.

Here are some notes on the output signal level and impedance of a condenser mic. Both signal level and output impedance are determined by the integrated buffer amplifier in the mic capsule. The **spec sheet for the Neumann TLM103 condenser mic** lists an output voltage sensitivity of 23 mV per Pascal of pressure at 1 kHz into a 1-kOhm load, and an output impedance of 50 Ohms. This 23 mV/Pa voltage sensitivity is quite a bit larger than the 1-3 mV/Pa sensitivities typically found in dynamic microphones. Likewise, the 50-Ohm output impedance is substantially lower than the 150-300 Ohm impedances commonly found in dynamic microphones.

Lastly, the Neumann TLM103 condenser mic has a very low self-noise level (7 dB-A) and a very high maximum sound pressure level (138 dB SPL). This means that the TLM103 has an extremely large dynamic range of ~131 dB, making it capable of capturing the softest sound or loudest sound without adding noise or distortion – just another feature making this microphone a mainstay in most music recording studios.

In the next chapter, we'll take a look at the topic of stereo microphone techniques.

B6. Microphone Techniques



[eyeliam, CC BY 2.0, via Wikimedia Commons](#)

There is much written in textbooks, user guides, online publications, and blogs about the science and art of using microphones for recording music in the studio. It is critically important for engineers and artists to choose the appropriate microphones and to position them correctly, because getting the “right sound” up front is absolutely necessary in the recording process. Two good references on this subject are:

Shure Inc., **Microphone Techniques for Recording**, 2014.

D.M. Huber and R.E. Runstein, **Modern Recording Techniques**, 9th ed., Taylor & Francis, 2018.

There is so much expert knowledge contained in these references – I highly recommend taking a look at them. In this chapter, I would like to mention a few fundamentals about microphone setups.

1. Close Mic Placement

Placing a directional microphone, such as a mic with a cardioid polar pattern, within three feet of an instrument is a good way to record a track that isolates this instrument. Not much sound from the ambient environment – other instruments and room reverberations – will be picked up by the mic. Having a recorded track that is mostly direct (just the one instrument) and dry (no reverb) gives you a lot of creative freedom in the subsequent effects and mixing processes.

The “leakage” of sound from another instrument or sound source can be minimized by using these tips:

- Better mic placement – closer to desired source; angled away from other sources
- Use sound barriers – “gobo” (go-between) provides isolation; isolation booth
- Adhere to a 5:1 rule – for every unit of distance between a mic and its source, a second mic and its source should be separated by at least five times that distance.

A note about “close mic’ing” – the position of a directional microphone, *i.e.*, its distance from the instrument as well as its lateral position, will have a significant effect on the tonal balance (timbre) of the sound picked up by the mic. The **Proximity Effect** discussed previously is a prime example of the microphone’s frequency response coloring the tonal balance as the mic moves closer to the source. So, artists and engineers have developed many positioning strategies for just about every musical instrument made. You can read about these in the two references listed above.

2. Distant Mic Placement

In this setup, the mic is positioned at a distance of 3 feet or more from the instrument or sound source. Often, a natural tone balance can be achieved by placing the mic at a distance that's on the order of the largest dimension of the instrument. At this distance, the full timbre of the instrument can be captured from the outgoing sound wave. Additionally, there will be sound captured from the room's acoustic reverberation, mixing in naturally with the direct sound signal to give an overall rich, wet sound.

Distant mic'ing is typically used with large instrumental ensembles, such as a symphony orchestra or chorus. The mic's distance is adjusted to strike an overall balance between the ensemble's direct sound and the space's reverberant sound. This technique gives a full, open feeling to the recorded sound. The one big 'caveat' here is that the acoustics of the room, hall, church, etc. become a permanent part of the recording – so the recording engineer and producer had better get this right during tracking – there's no "fix in the mix" here.

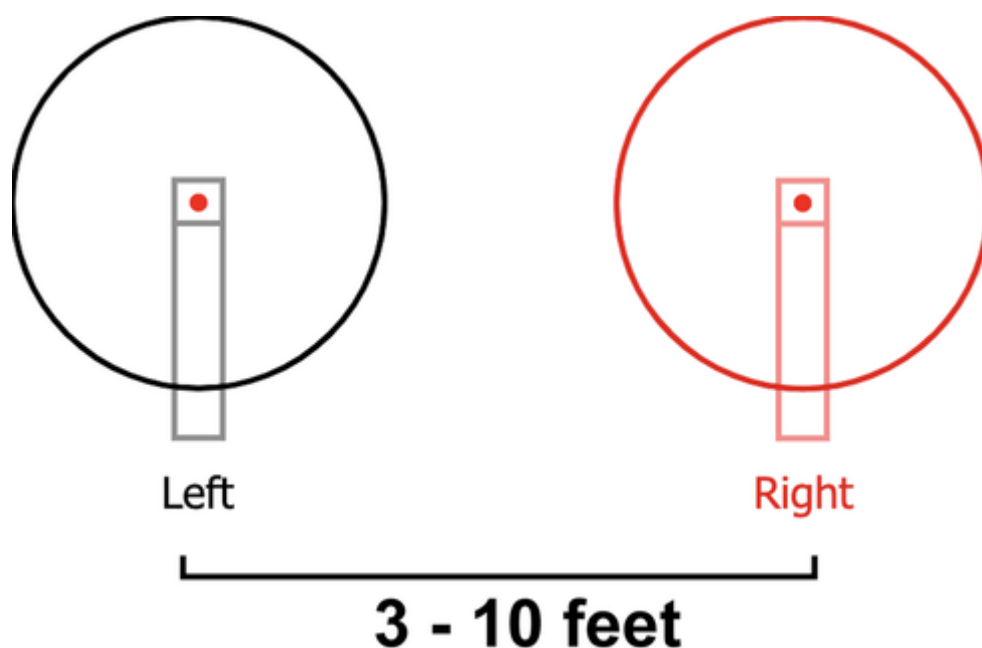
3. Ambient Mic Placement

As you have by now guessed, this third microphone placement is at such distance that it picks up just the room's reverberant sound. Due to the inverse square law of a soundwave's amplitude versus distance traveled, the direct wave from the source is smaller in amplitude at the microphone than the combined amplitude of the huge number of waves reflected from the room's enclosing surfaces. In the studio, ambient microphones are used to add a sense of space or natural acoustics back into the sound. The ambient sound is recorded to separate tracks and blended in small amounts into the music during the mixing process.

Stereo Microphone Techniques

Stereo mic'ing typically uses two microphones to capture a coherent stereo image, much like your two ears do when listening to a musical performance on stage. The two output signals from the microphones are sent to a stereo track in the digital audio workstation or to two channel strips, one panned hard left and one panned hard right, on a mixer console. A stereo mic pair can be used in either close, distant, or ambient placements, and can be used for single instruments, vocals, ensembles of all sizes, or large stage productions such as symphonic orchestras and chorus. There are several stereo mic'ing techniques in widespread use today. Three of the more popular techniques are outlined below.

1. Spaced Pair (sometimes called “A/B”)



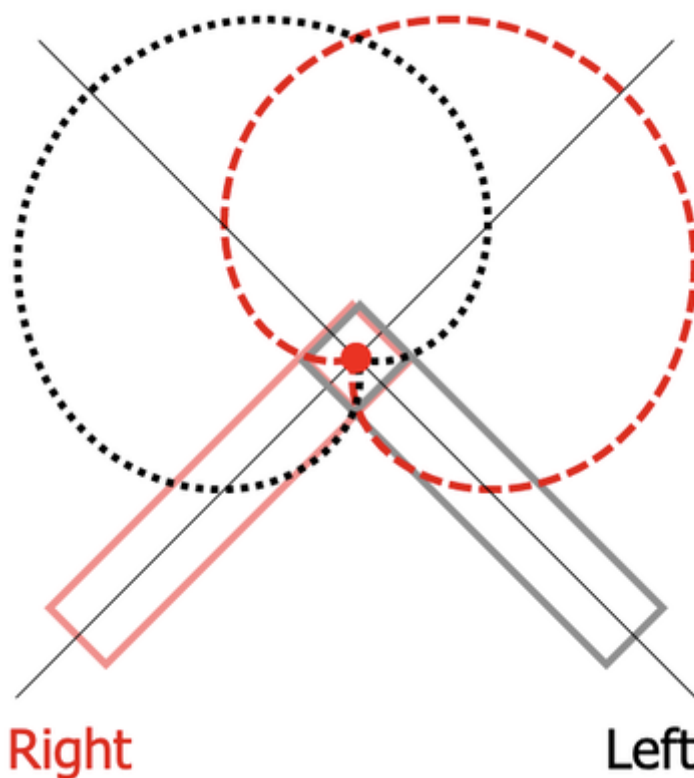
The spaced pair technique employs two cardioid or omni directional microphones spaced 3 – 10 feet apart from each other to capture the stereo image.

In a previous post, we discussed the psychoacoustics of **sound source localization**. The ability of the human auditory system (two ears and brain !) to identify the direction from which a sound is emanating is called sound source localization. Humans can locate this sound source in space with extreme precision – within 2 degrees in the horizontal plane. This remarkable feat is accomplished by the brain's ability to process the binaural signal coming from the ears.

Neuroscientists believe that sound localization in the horizontal plane relies on two “cues”: the sound amplitude (loudness) difference between the two ears (inter-aural level difference ILD), and the time difference (delay) of sound reaching each ear (inter-aural time difference ITD). The brain uses both cues to localize sound sources.

Pretty much the same thing is happening here – the two microphones, like your ears, map out the stereo field by using the amplitude difference and time difference cues. Because of the relatively wide spacing between the microphones, a high “resolution” of the sources of sound across a wide stereo field is possible. Unfortunately, a drawback of the wide spacing between the microphones is the strong potential for phase cancellations between the left and right channels, due to differences in a soundwave’s arrival time at one mic relative to the other. In a mono playback of the stereo track, these phase differences may lead to certain frequencies dropping out of the sound.

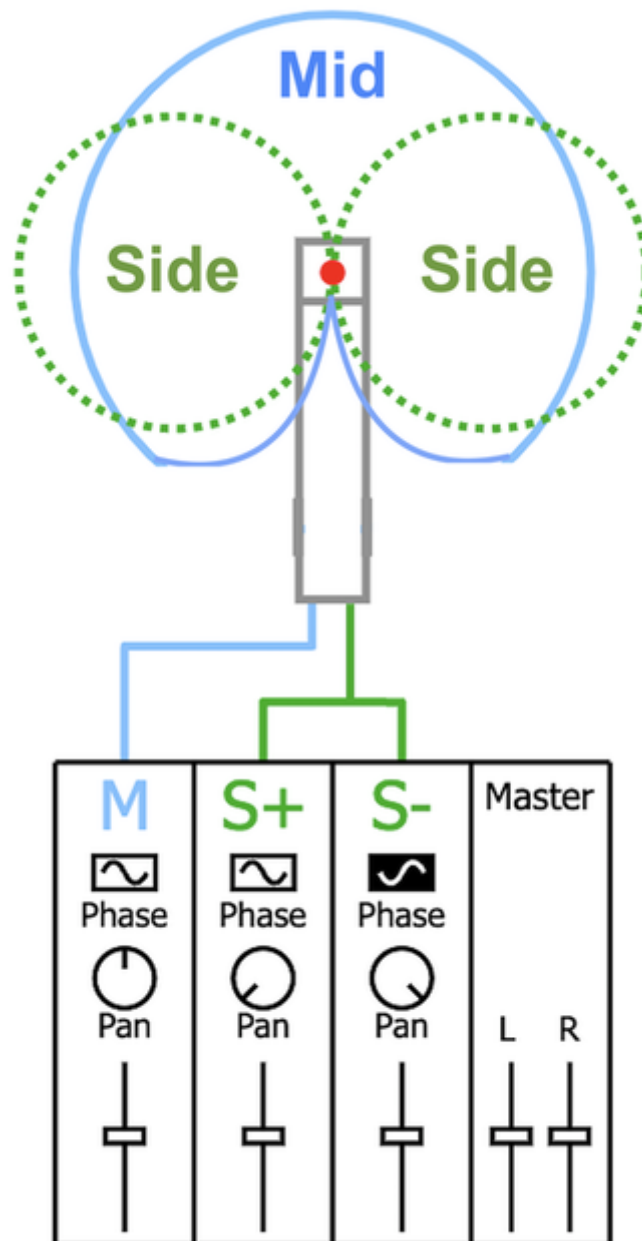
2. Coincident Pair (sometimes called “X/Y”)



[Iainf 23:51, 21 September 2007 \(UTC\), CC BY-SA 3.0, via Wikimedia Commons](#)

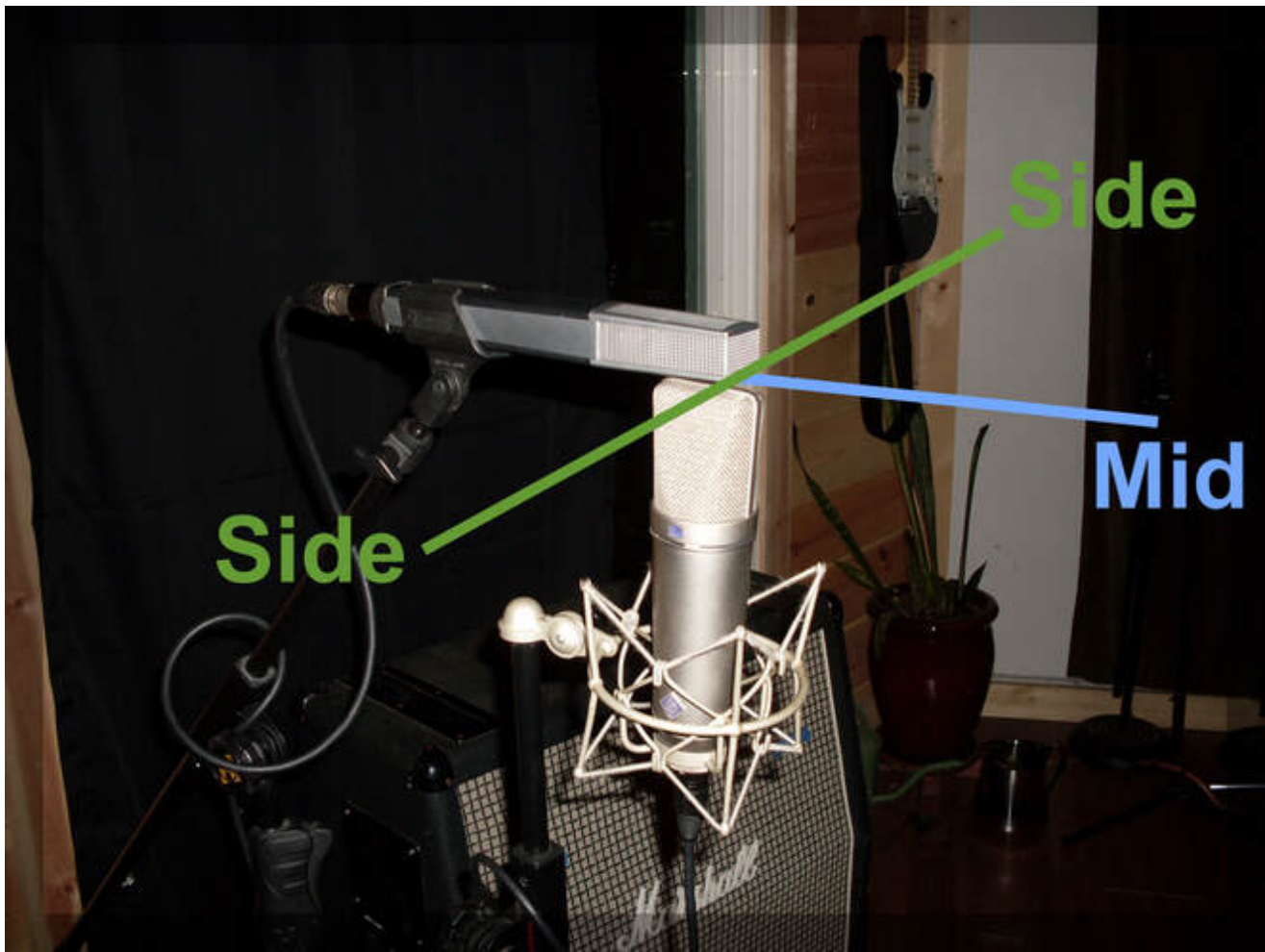
The coincident pair technique employs two cardioid microphones of the same type and manufacture with the two mic capsules placed as close as possible to each other and facing each other at an angle ranging from 90 – 135 degrees (depending on the size of the sound source and the desired stereo field width). Sound waves arrive at both microphones at the same time, thereby eliminating any potential phasing problems such as occurred in the A/B technique above. The drawback here is that the stereo information comes from the directionality properties of the two microphones that create only amplitude difference cues. The width of the stereo field is generally good, but may be limited if the sound source is very wide.

3. Mid-Side (M/S)



Adapted from [Iainf 23:51](#), 21 September 2007 (UTC), CC BY-SA 3.0, via [Wikimedia Commons](#)

The Mid-Side technique employs two coincident microphones, one with a cardioid pattern and the other with a bi-directional (figure-8) pattern. The cardioid mic (“Mid”) faces directly at the center of the sound source and picks up primarily on-axis sound. The bi-directional mic (“Side”) faces left and right, and picks up off-axis sound.



Adapted from Roadside Guitars, CC BY-SA 2.0, via Wikimedia Commons

The Mid and Side mic signals are recorded to two separate tracks. In Section C. Mixing, we will encounter the topic of **Mid-Side processing** to widen (or narrow) the stereo image. A stereo imager plug-in is used to adjust the level of the Side channel relative to the level of the Mid channel. By increasing the gain (positive dB) of the Side channel, it is shown by some straightforward algebra that the width of the stereo field is widened. Conversely, by decreasing the gain (negative dB) of the Side channel, the width of the stereo field is narrowed. The stereo output Left (L) and Right (R) channels are created by a mathematical 'transform' of the Mid (M) and Side (S) channels inside the imager plug-in,

$$L = (M + S)$$

$$R = (M - S)$$

Mid-Side stereo mic'ing is completely mono-compatible, and is widely used in broadcast and film applications.

B7. Low-Latency Monitoring



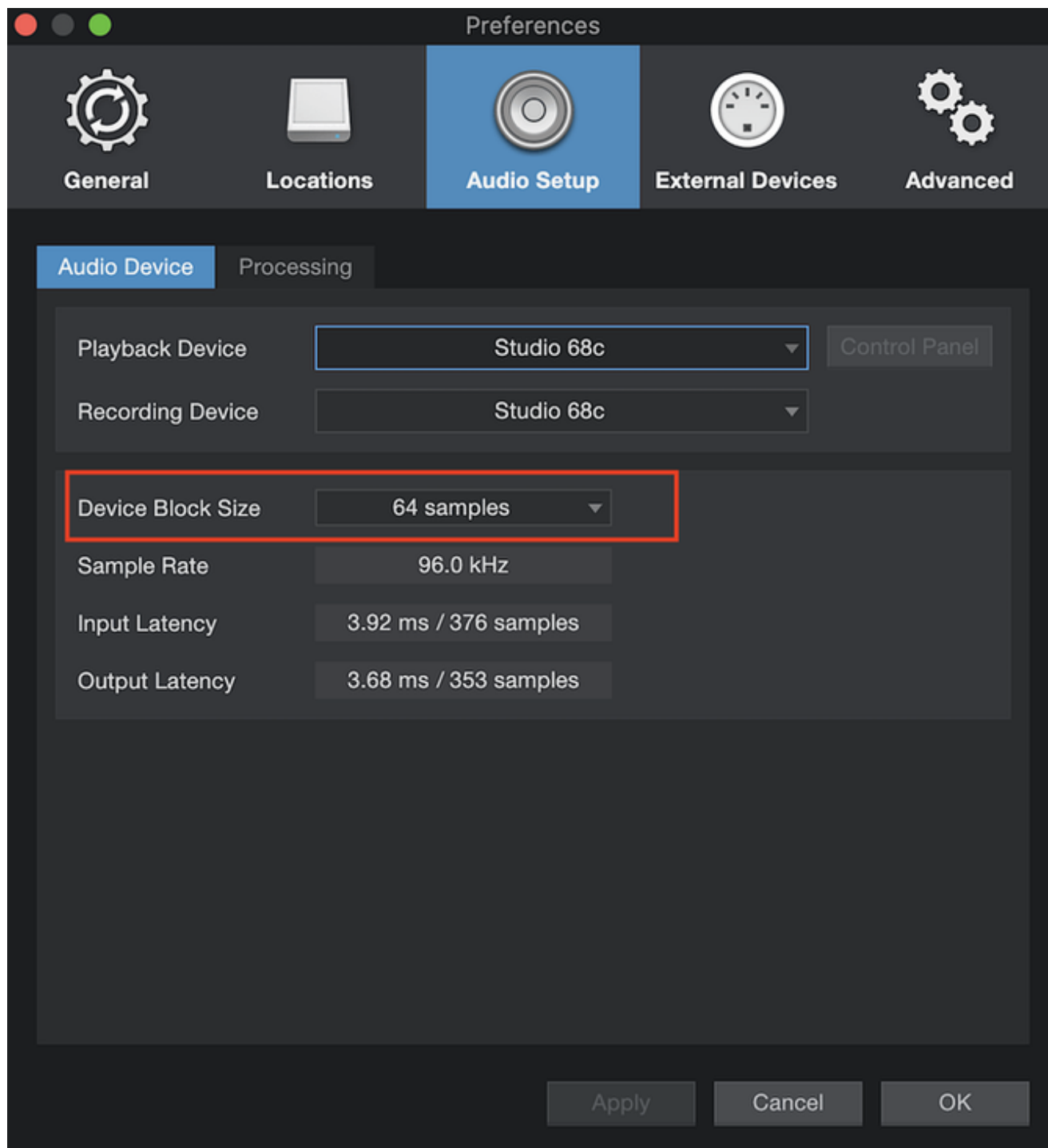
"Guitar girl" by Dennis AB is licensed under CC BY-NC-ND 2.0.

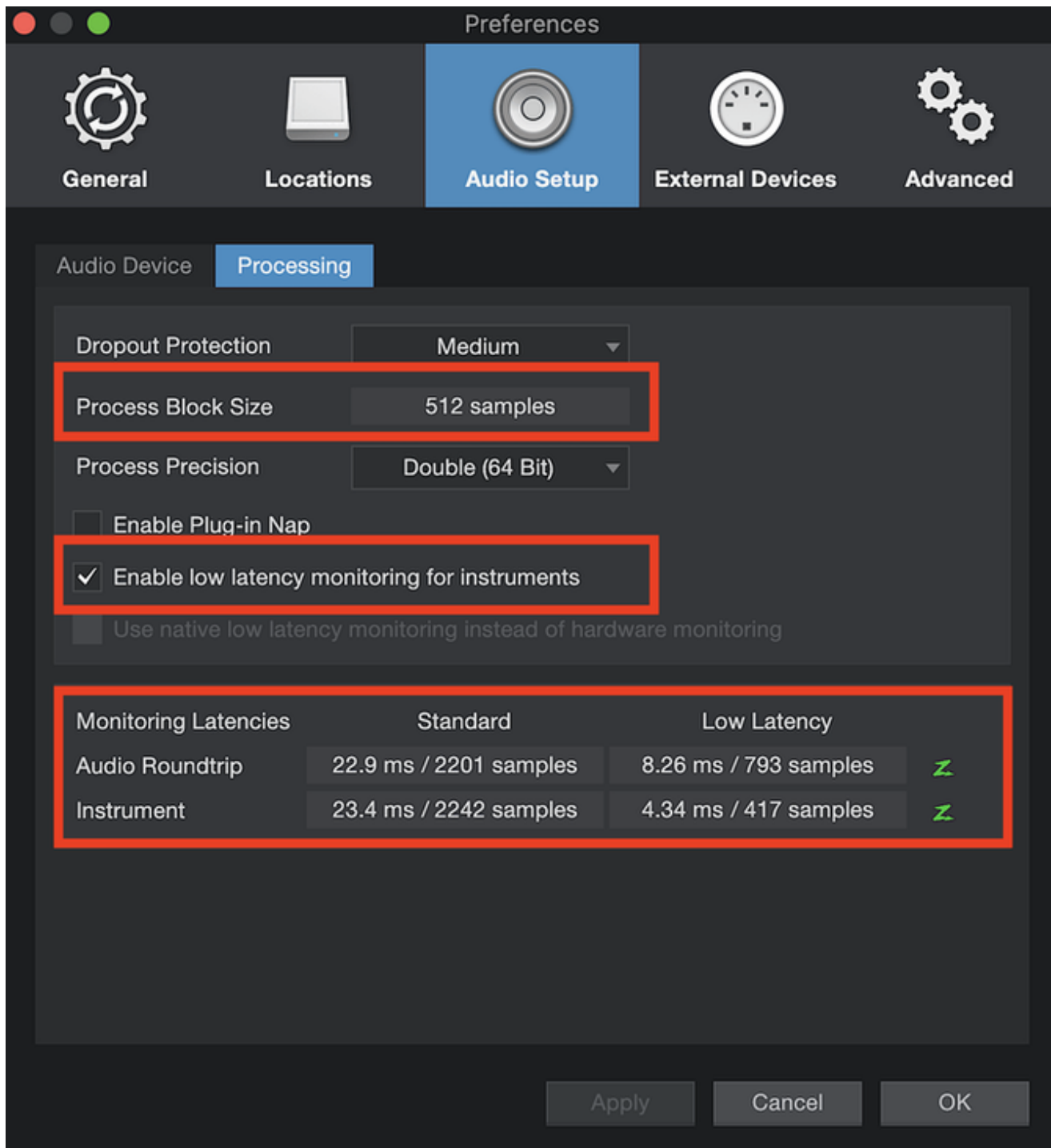
In an earlier chapter on **Latency**, we looked at the buildup of time delays, measured in milliseconds (ms), of digital audio signals as they pass through the hardware/software of the computer-based recording system. There are two activities associated with recording when latency can be a significant issue – monitoring and overdubbing. Let's review the monitoring process in this chapter. We can then use this information to create what's called a **Cue Mix**, a topic that we'll address in the next chapter.

When you are working with a large number of audio tracks and virtual instruments, there is a need for buffering data in the computer so that all the CPU computations required for signal processing in the digital audio workstation (DAW) software can be performed accurately and "on time". You can increase the buffer size to cope with all these computations, but this traditionally comes at the cost of greater latency (delay) when monitoring audio inputs or playing virtual instruments. Set the buffer size too low, and audio dropouts and glitches will occur.

In the **PreSonus Studio One DAW**, the tasks of (1) audio playback with all effects plug-ins active and (2) monitoring of audio inputs and virtual instruments are handled as separate processes. This, in effect, lets you use a large processing buffer to handle the computationally-intensive audio playback and effects processing tasks, while keeping latency low for audio input and virtual instrument monitoring.

The latency that you hear when monitoring audio inputs or playing virtual instruments is based primarily on the Device Block Size that you specify in the Audio Device setup window (see screen capture below). For the lowest latency, Device Block Size should be set to the lowest setting that maintains adequate CPU performance. The Audio Dropout Protection system uses its own buffer for playback and processing of audio and instrument tracks, distinct from the Device Block Size setting. The size of this buffer (called the Process Block Size) depends on the Dropout Protection level that you specify in the Audio Processing setup window (see screen capture below). If you use Native Low-Latency Monitoring, the Dropout Protection level has no effect on audible latency. As long as the Process Block Size is larger than the Device Block Size, you have the option to use Native Low-Latency Monitoring in the PreSonus Studio One DAW software. Here are the Device Block Size and Process Block Size that I typically use for my recording needs:

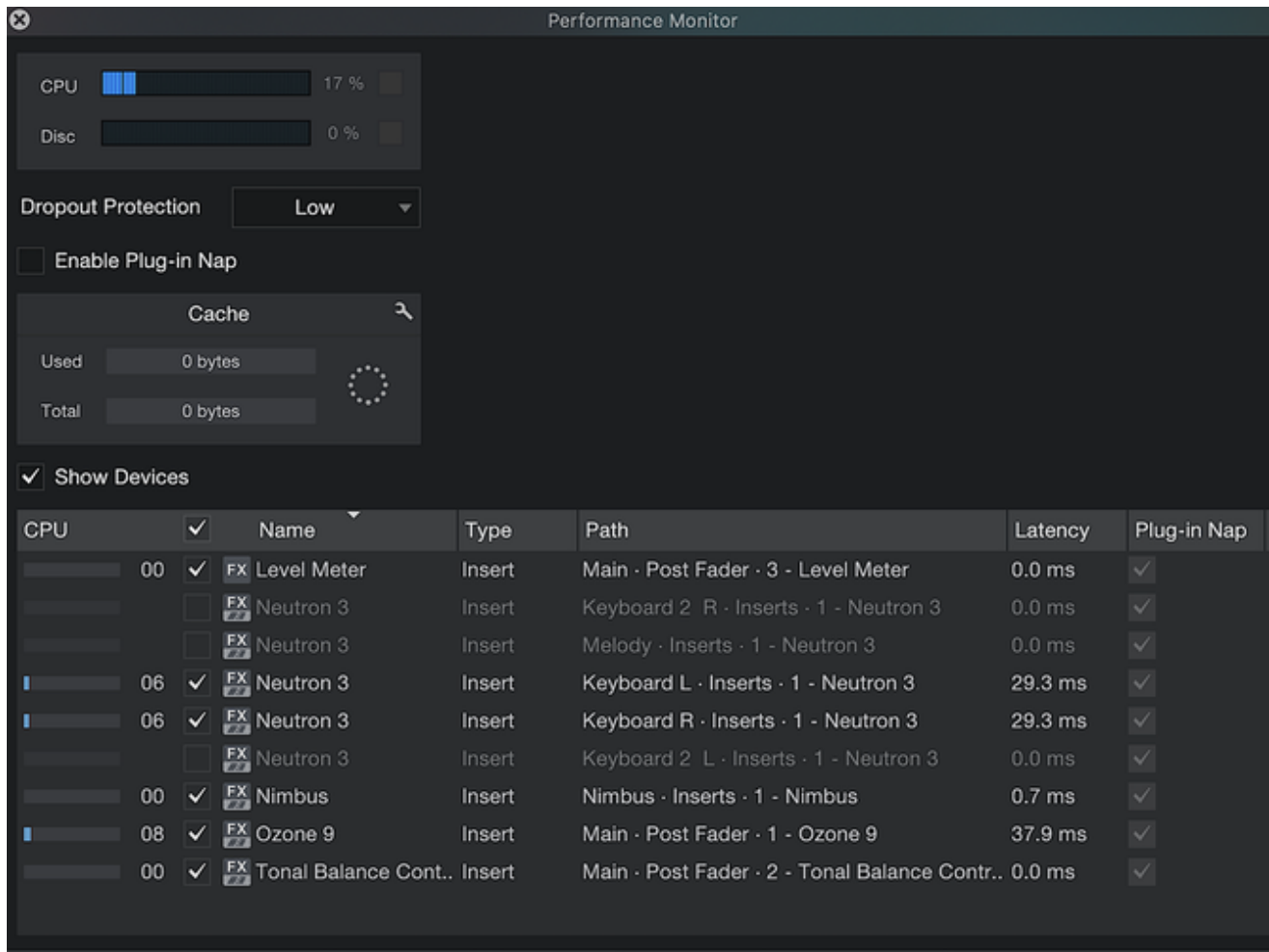




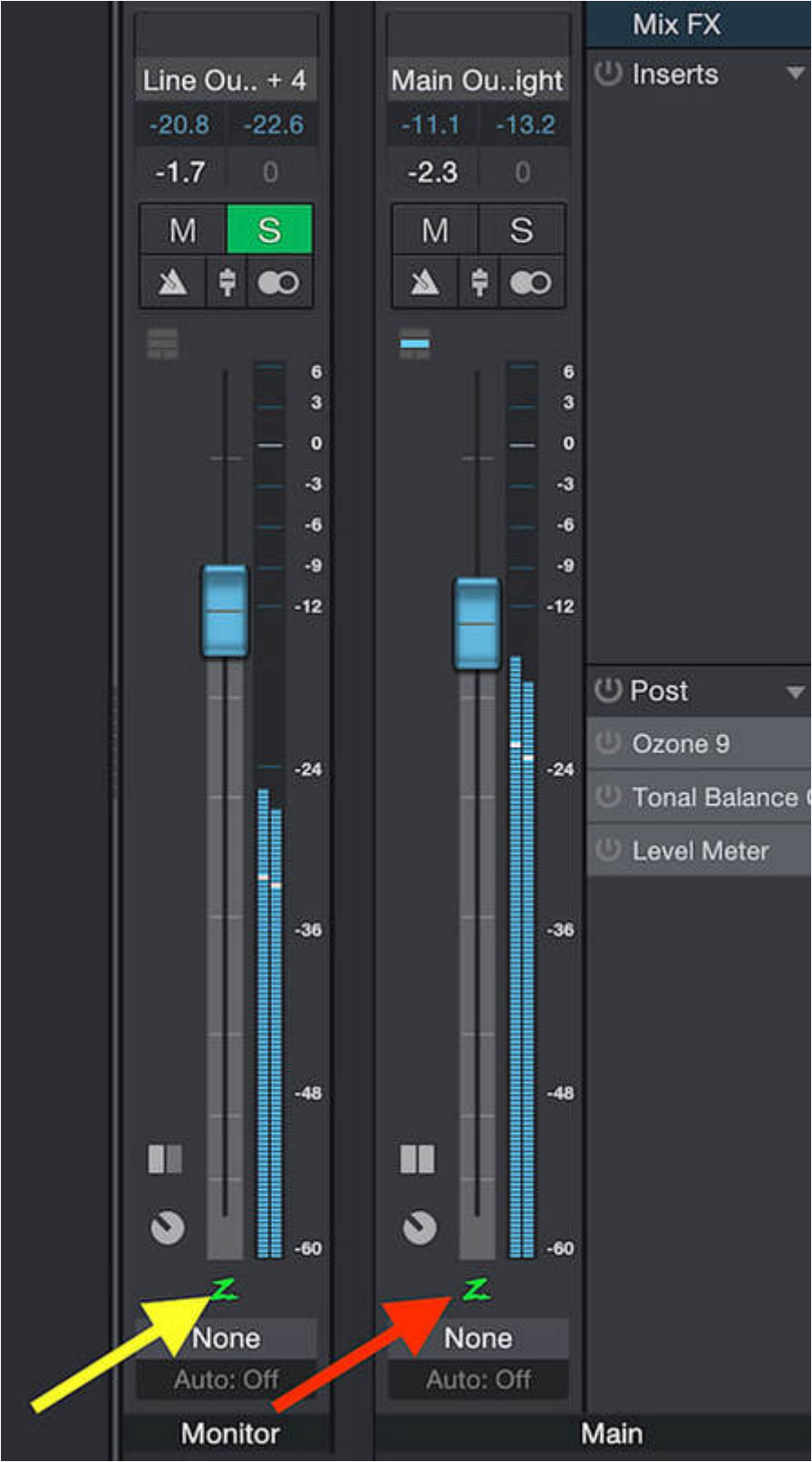
For low-latency monitoring when playing virtual instruments, click the "Enable low latency monitoring for instruments" box.

The Monitoring Latencies display shows you the latency values for audio inputs (round-trip, from input to output) and virtual instruments, based on the current Device Block Size and Dropout Protection settings. The "Standard" column shows the latency for the current settings if you choose not to use Low-Latency Monitoring, while the "Low Latency" column shows the latency for the Native Low-Latency Monitoring system.

When monitoring audio inputs and/or virtual instruments through the Native Low-Latency Monitoring system, any inserted effects (FX) on the corresponding Channel continue to function and can be heard in real time, provided that they add 3 ms or less of latency. Any inserted plug-ins that introduce more than 3 ms of latency are not audible in the monitoring path while a Channel is armed for monitoring or recording. Below is a screen capture of the CPU performance monitor, showing the latencies of several different FX plug-ins.



With the Native Low-Latency Monitoring system configured, you can toggle low-latency monitoring on and off for the Main Mix output and **Cue Mix** (see next chapter) output, by clicking the Enable Low-Latency Monitoring button ("Z" symbol) below the volume faders for the two mix outputs. When Native Low-Latency Monitoring is enabled, the "Z" button turns green in color.



When recording, you want to hear your performance as you play, as well as the performances of others if you're playing with a group. When monitoring the music being played through your computer's signal path, latency will be experienced as short delays between the time you play a note and the time you hear it on your headphones. If this delay is excessive, say, greater than 10 ms, then it will be nearly impossible for you to play well on your own instrument, not to mention playing with a group. If all of your audio inputs are being recorded to audio tracks (*i.e.*, you are not using any virtual instruments whose sound originates inside the computer software), then it is better to use the "zero latency" option afforded by external monitoring in hardware. This external hardware monitoring is accomplished through front-end mixers built into many audio interface units or through a stand-alone mixer console. If you must use "in the box" software monitoring, primarily because you are playing virtual instruments, then you most likely will want to enable the Low-Latency Monitoring feature. In the example above, standard monitoring latencies exceed 20 ms. This is unacceptable. The Low-Latency Monitoring option gives values less than 10 ms. This is acceptable.

Setting up a low-latency Cue Mix Monitor for your musical performers will be outlined in the next chapter.

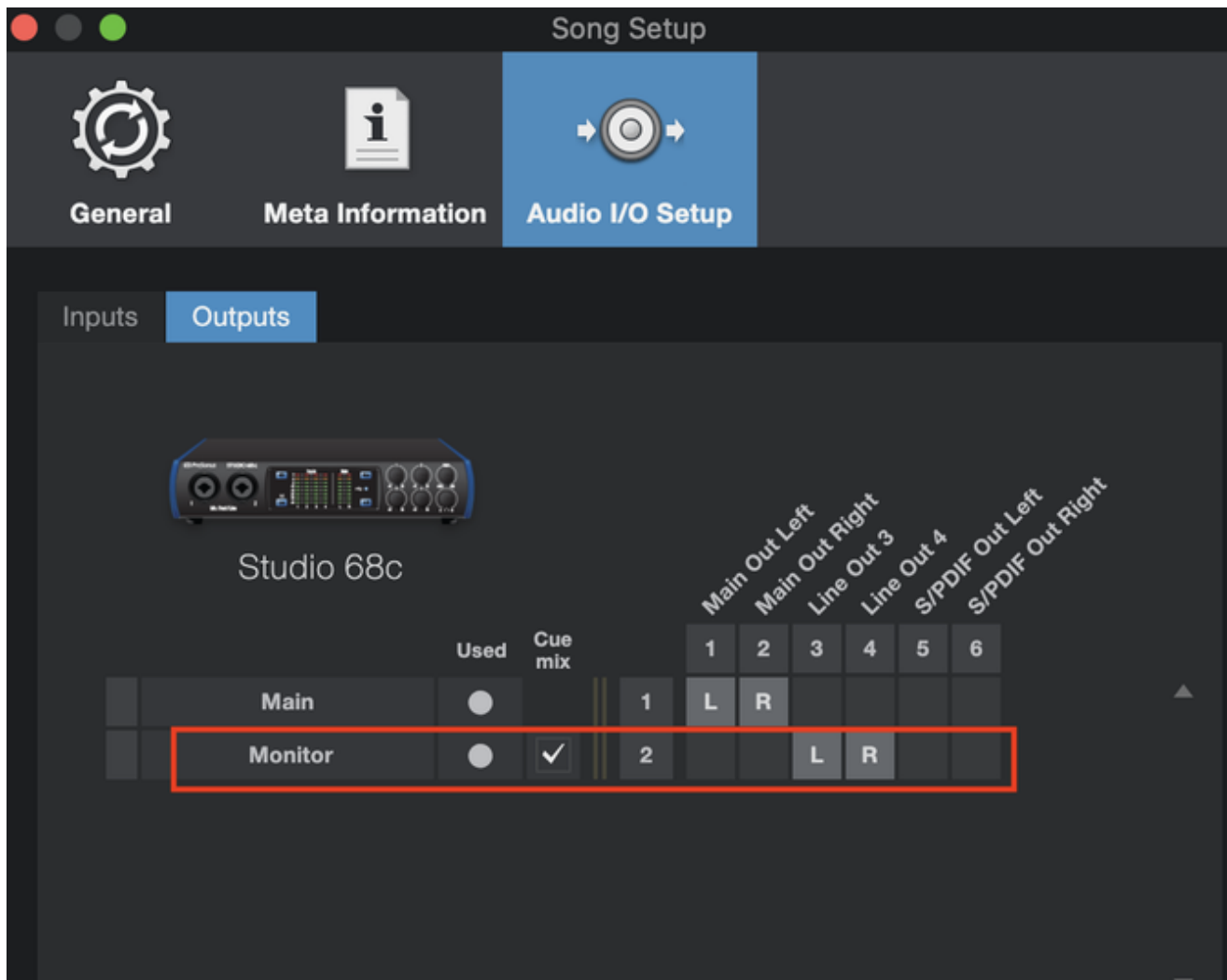
B8. Cue Mix Monitoring



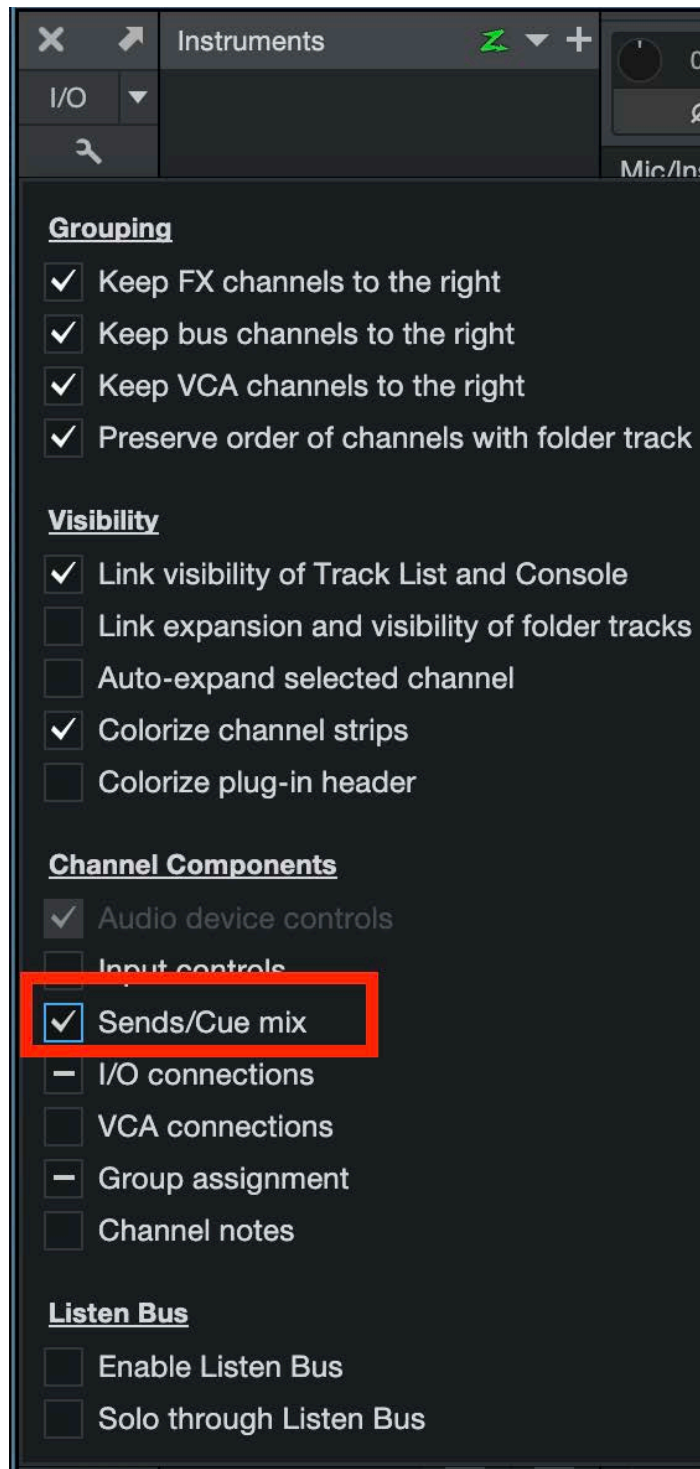
Jeff Wilson, CC BY 2.0, via Wikimedia Commons

In the last chapter, we looked at **low-latency monitoring** for audio input and virtual instrument recording. Now, we can use that low-latency monitoring to set up what is called a **Cue Mix** output – a mix that is separate from the Main Mix output and is provided to the musicians in their headphones during the recording session. For example, when recording vocals, the engineer and vocalist often need to hear different mixes. Many vocalists want to hear their vocal boosted in the mix, possibly with some reverb to make it sound natural, while the engineer will want to focus on how the performance balances with the rest of the mix. We can set up a Cue Mix for the vocalist, separate from the Main Mix for the engineer.

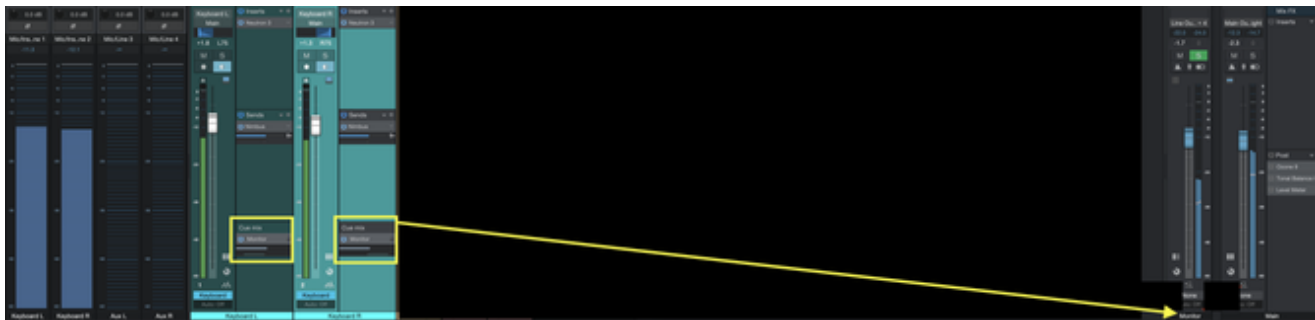
The first step in building a Cue Mix is to create an additional Output Channel. To do this, open the Audio Output Setup window in your Digital Audio Workstation (DAW) and add a new Stereo Output Channel – call it something like “Monitor”. Next, specify that this Monitor output is a Cue Mix output by clicking on the Monitor’s Cue Mix check box. You can create as many Cue Mixes as your audio interface has available stereo outputs. Here’s the Audio Output Setup window in my **PreSonus Studio One DAW**.



Next, create Cue Mix Sends in your console channels by checking the corresponding box in the channel components list (under the "I/O" tab in the PreSonus Studio One mixer console).



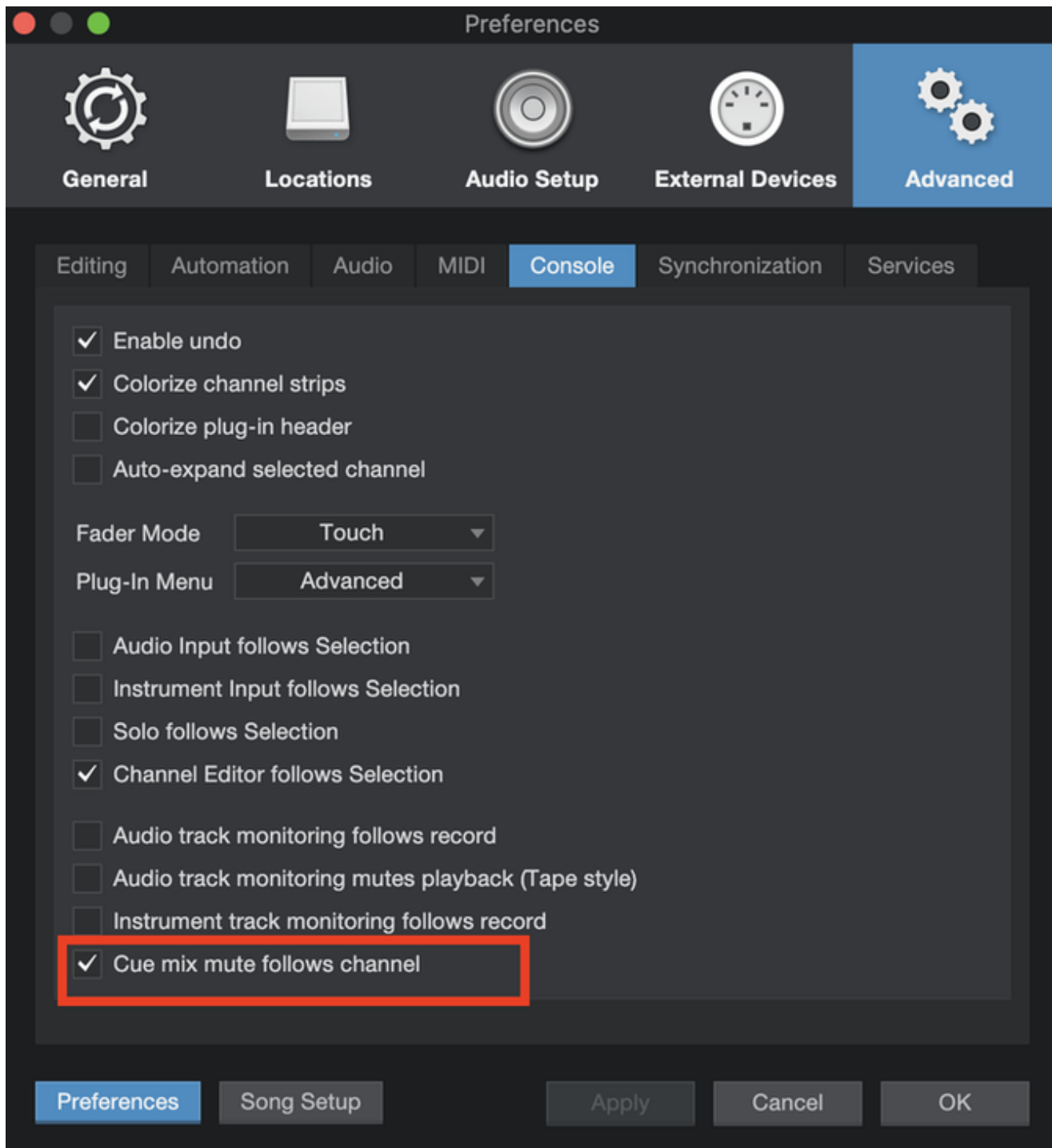
Now, you have Cue Mix Sends in your channels. There are level faders and pan controls for each of these Cue Mix Sends, so you can adjust the Cue Mix in just the way your musicians want to hear it



Standard Monitoring Cue Mix (click image to enlarge)

Let's return to our example of recording live vocals. For a vocalist to be comfortable and perform well, it is important that the performance sound as natural and as polished as possible. The vocalist needs to hear herself prominently in the mix, with no audible delay of her voice. And blending in some reverb provides a little ambiance so that her voice is not dry and lifeless.

To be able to make Cue Mix Sends from Bus/FX Channels, there is a 'quirk' in the PreSonus Studio One DAW – you have to check a box in the Preferences/Advanced/Console window.....



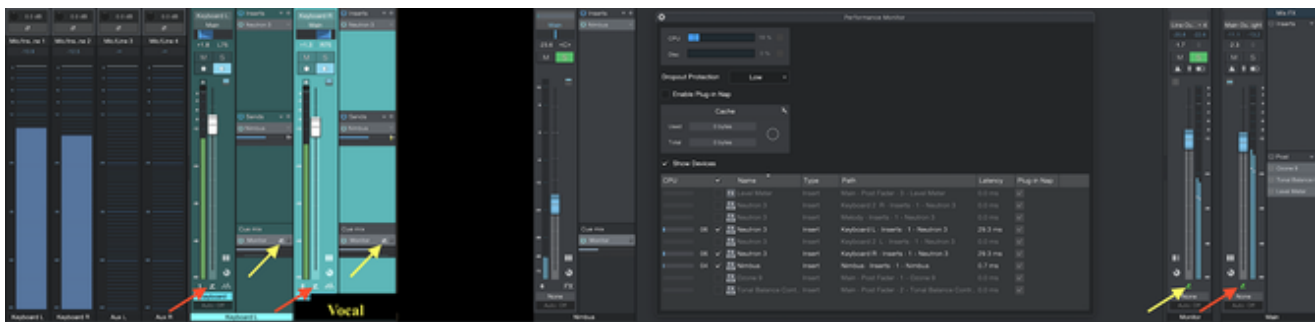
Now, you can create a pre-fader Send on the Vocal Channel to an FX Channel with your favorite reverb effect. And from the Cue Mix Send on the FX Channel, a small amount of the wet (100%) vocal reverb signal is sent to and mixed into the Monitor output. (The dry vocal signal is sent to and mixed into the Monitor output from the Cue Mix Send on the Vocal Channel.) You can probably get a better picture of this by following the signal paths in this screen capture of the mixer console



Reverb in Vocal Cue Mix Monitor (click image to enlarge)

Lastly, engage the Enable Low-Latency Monitoring ("Z") button below the level fader on the Cue Mix Monitor output being used by the vocal channel. (Note: Channels that are able to be monitored using Native Low-Latency Monitoring display a "Z" mark at the bottom of their channel strip.) The vocalist now hears the live low-latency input, as well as the rest of the cue mix, including the reverb effect. Adjust the level of the vocal and other Channels in the Cue Mix to the vocalist's liking, and we're ready to record.

Here's what the mixer console looks like – enabled for low-latency monitoring of the cue mix.



Low-Latency Cue Mix Monitoring (click image to enlarge)

Finally, I want to reiterate these words from the previous chapter

"If all of your audio inputs are being recorded to audio tracks (i.e., you are not using any virtual instruments whose sound originates inside the computer software), then it is better to use the "zero latency" option afforded by external monitoring in hardware. This external hardware monitoring is accomplished through front-end mixers built into many audio interface units or through a stand-alone mixer console. If you must use "in the box" software monitoring, primarily because you are playing virtual instruments, then you will want to enable the Low-Latency Monitoring feature."

By setting up a low-latency Cue Mix Monitor as discussed in this post, we can ensure that musicians hear their instruments with low latency, in a custom mix that can include effects. Simultaneously, we can listen to a completely independent main mix, allowing us to focus on sound engineering while the artists focus on the musical performance.

B9. Driver Error Compensation



In this chapter, I'll review the important topic of Driver Error Compensation in the recording process. Click on the image above to watch the amazing YouTube video on this subject. It was made by Gregor Beyerle of PreSonus. While he specifically addresses the **PreSonus Studio One** digital audio workstation (DAW), this topic is universally applicable to all DAWs.

During overdubbing in the recording process, errors in aligning tracks on the time axis will inevitably occur. These errors are caused primarily due to latency issues. In a previous chapter, I provided an overview of **what is latency** and how does it affect the two recording activities of monitoring and overdubbing.

Latency refers to the buildup of time delays, measured in milliseconds (ms), as digital audio signals pass through the hardware/software of the computer-based recording system. There are four basic contributors to latency:

1. the analog-to-digital and digital-to-analog signal conversions in the DAC circuit of the audio interface,
2. the data transfer speeds of the USB or Thunderbolt buses between the audio interface and the computer,
3. signal processing speed of the digital audio workstation (DAW) software running on the computer CPU, and
4. data input/output speed of flash memory storage (solid-state drives (SSD)). The audio files are most likely kept on an external SSD, so a Thunderbolt-3 connection to the computer is highly desirable.

The issues of latency affecting the monitoring process during recording were discussed in quite a bit of detail in the two chapters, **Low-Latency Monitoring** and **Cue Mix Monitoring**. In this chapter, we'll tackle the issues of latency affecting the overdubbing process during recording.

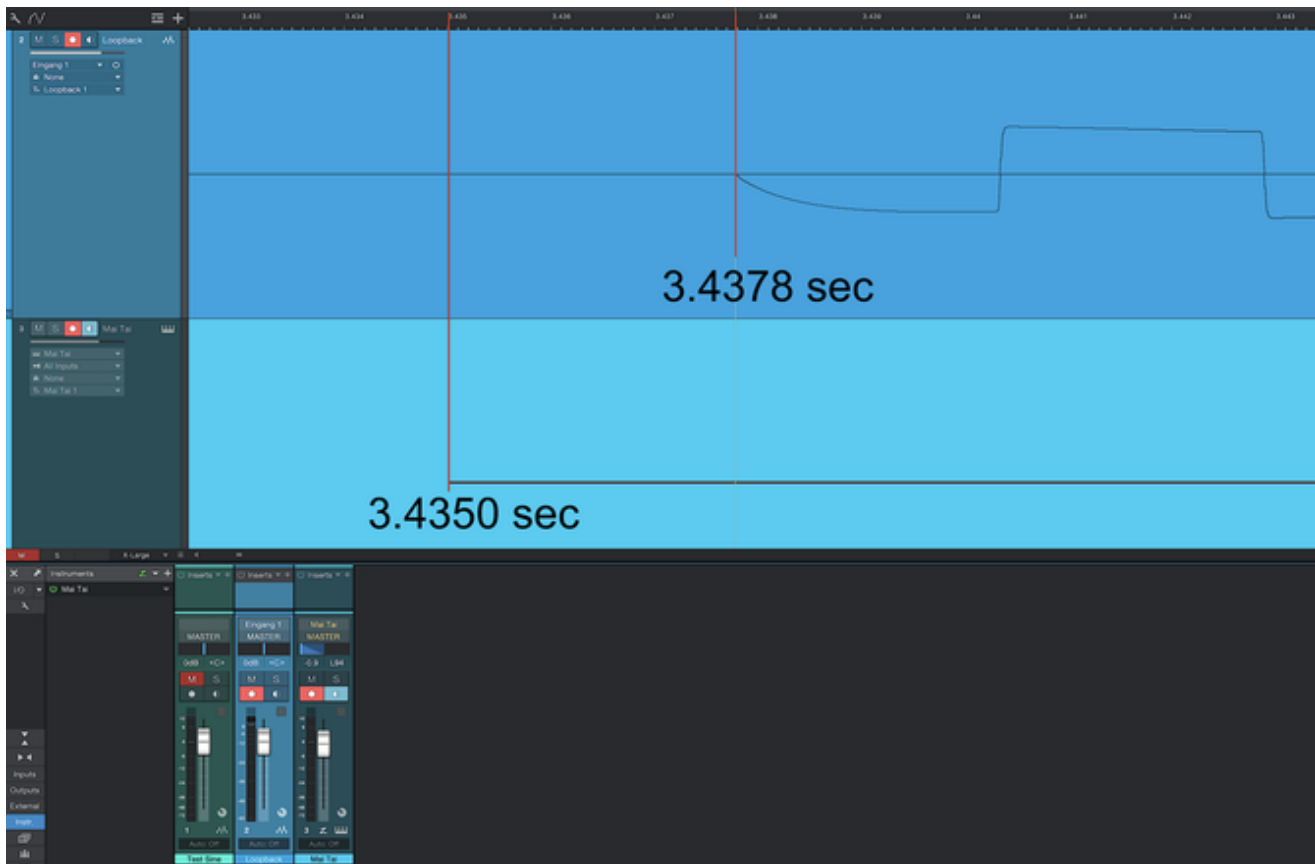
Overdubbing is the practice of listening to the playback of previously recorded tracks and recording an additional, separate track to the computer audio file that is “in synch” with the previously recorded tracks. The key word here is “in synch” -- timing is everything. Envision the signal flow -- a playback signal originates from the audio file on an external SSD drive and travels all the way to the monitor output on the audio interface. Upon hearing this signal, you play right along with it, hopefully “in synch” . Your new signal at the audio interface input now travels all the way back to the external SSD drive and is recorded on a new track in the audio file. You can imagine that the round-trip latency will “place” the new track data somewhat delayed on the time line from the original playback track – *i.e.*, NOT in synch !

This is where the topic of “Driver Error Compensation” comes in. Driver error compensation is the official name for the correction that your DAW makes to try to overcome this round-trip latency by properly aligning the pre-recorded and newly-recorded tracks on the time axis. A process for “calibrating” your DAW’s driver error compensation is what is presented in the PreSonus video above. This calibration process is called a “loopback test”. This video generated MANY comments, wanting more information on how this works and wondering why such an important topic is buried so deep in the menus of the PreSonus Studio One DAW .

The basic operation is that the DAW “places” the newly-recorded track “backward in time” by an amount estimated by a calculation of the round-trip latency. The “loopback test” provides a correction to this time shift. These corrections (in number of samples (audio) or in milliseconds (MIDI)) may be positive or negative values, depending on whether the DAW overestimated or underestimated the round-trip latency. An example of aligning two audio tracks using driver error compensation was shown back in the earlier chapter on **latency**.

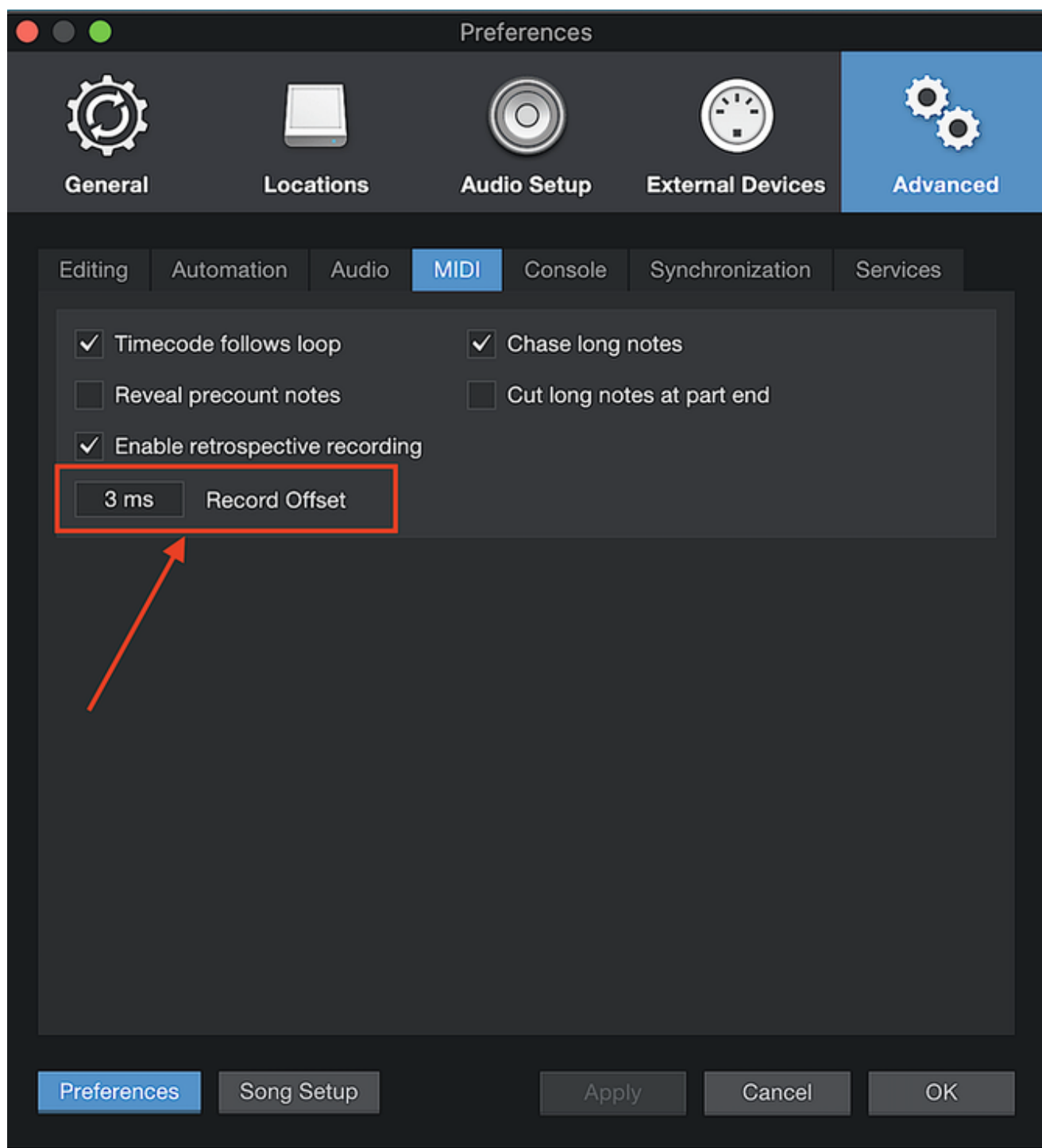
Here, I want to go a step further and bring into play instrument tracks. Multi-tracking of MIDI instruments is probably the most popular activity done in home music recording studios. MIDI is an acronym that stands for **Musical Instrument Digital Interface**. It's a way to connect devices that make and control sound —such as synthesizers, samplers, and virtual instruments (software). In many home studios, virtual instrument (VI) plug-ins populate a large number of instrument tracks in the DAW. By overdubbing, you can record each instrument track separately, and build up the entire sound that you want -- a symphony orchestra if you're patient and talented enough !

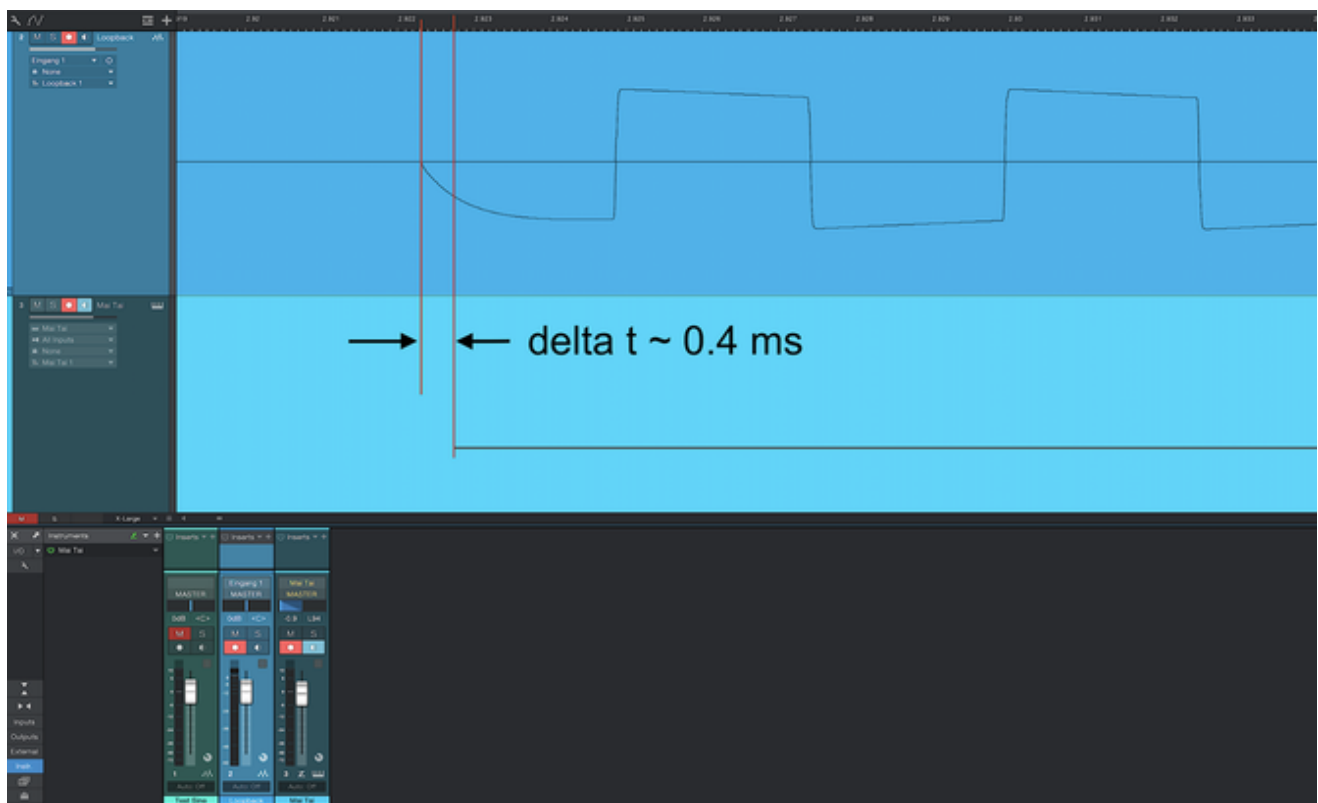
So, in order to record all the MIDI tracks “in synch”, we can go through another “loopback test” . Here, the virtual instrument Mai Tai (synthesizer software plug-in) is recorded on an instrument track, and the monitored sound produced in the DAW is fed back to an audio interface input (via a short audio cable) and recorded on an audio track. As shown in the figure below, the alignment of the MIDI notes and the audio waveforms is pretty good on a course time scale.



[Click image to enlarge](#)

By typing a 3 ms value (DAW only accepts integer values) in the MIDI Record Offset box in the Advanced Settings of the PreSonus Studio One DAW Preferences menu, we can achieve an even better alignment between the MIDI note and the audio waveform, as seen here.

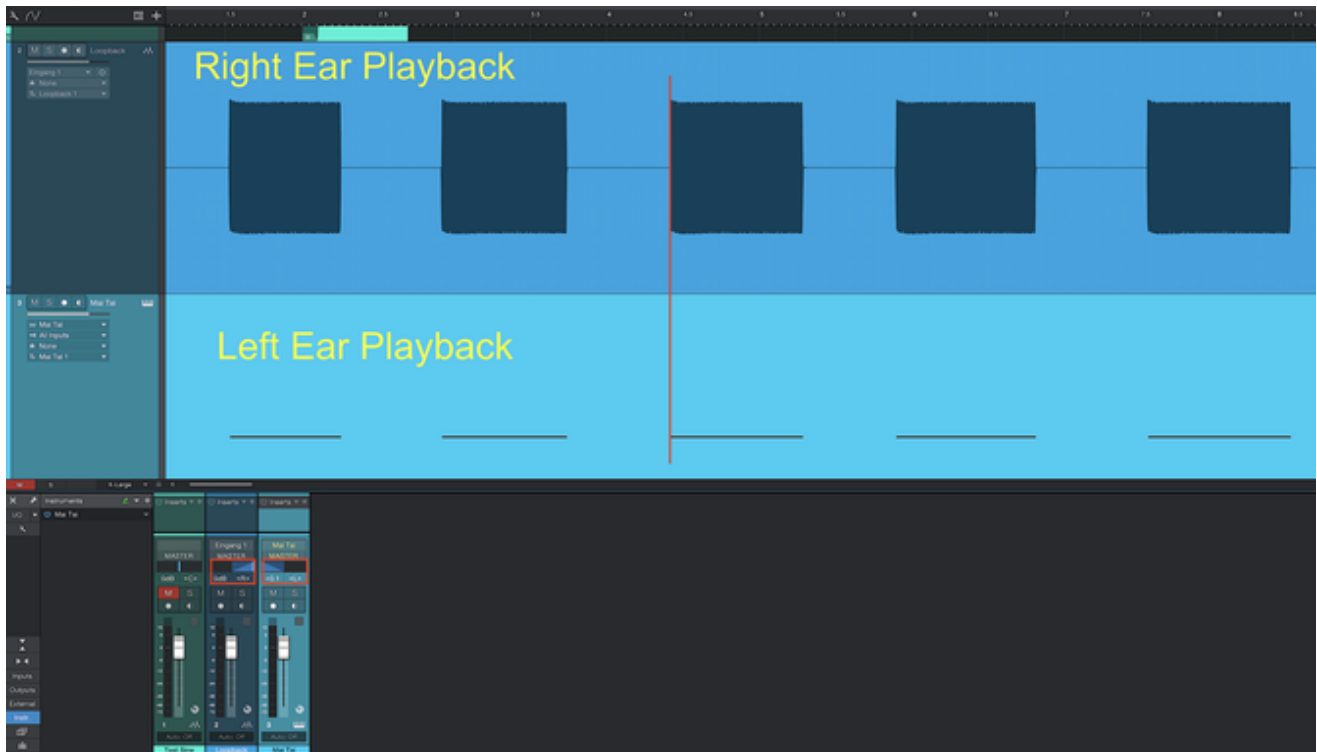




[Click image to enlarge](#)

The offset time resolution of the DAW is 1 ms, so that means that the alignment should be < 1 ms, but will never be 0 ms.

To listen to this alignment of notes in real time, I've played back the two recorded tracks, with the audio track (loopback channel) hard-panned right, and the instrument track (Mai Tai channel) hard-panned left. Listen carefully on earphones -- it's virtually impossible for the human auditory system to hear a time separation < 1 ms between the onset of the two notes, one in each ear. It just sounds like one note that is center panned, which is the objective of this driver error compensation exercise.



Click on image to enlarge



[MIDI Offset Correction](#)

click player to hear the playback of the tracks above

A Final Note:

The Record Offset value obtained from a loopback test is dependent on the settings used for your recording -- such as the sample rate, bit depth, and computation precision -- and is particularly dependent on the buffer sizes (Device Block Size and Process Block Size). Also, any FX plug-ins inserted in the audio playback channels should be disabled for this test and not used in the overdub recording. Different Virtual Instrument (VI) plug-ins will certainly have different latencies, too. Hopefully, these different VI plug-in latencies do not vary more than a couple of milliseconds, so that any resulting note misalignments remain undetectable. (Time quantizing on instrument tracks, a subject for future consideration, can help here.) Finally, using a different audio interface (with different I/O drivers) will definitely affect latency. So, bottom line -- if you change any of these things, a loopback test should probably be run again to get a more accurate value for the Record Offset.

B10. MIDI Recording

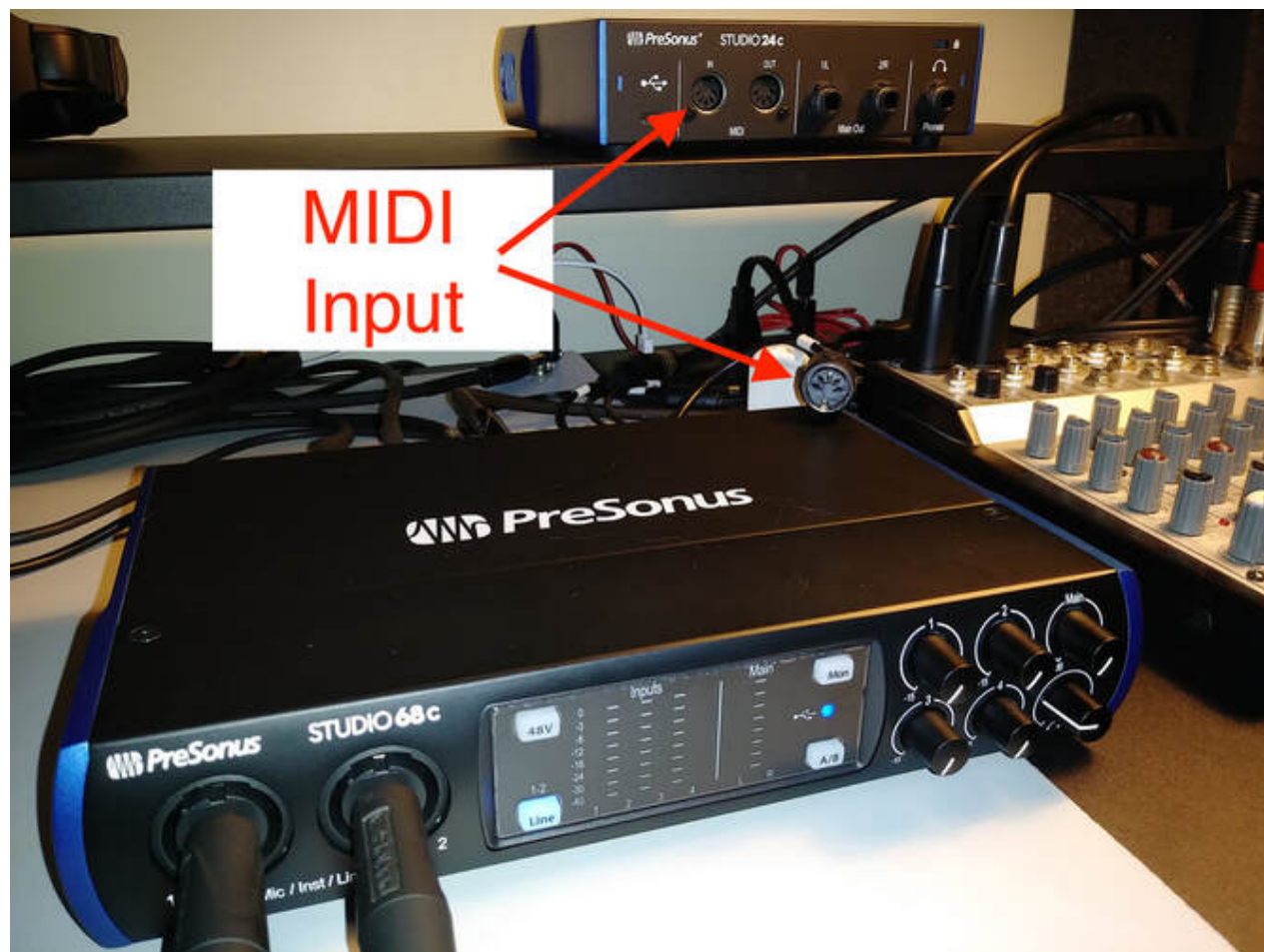


In this chapter, I want to show how to set up the music studio to record virtual instrument tracks. Multi-tracking of MIDI instruments is one of the most popular activities done in home music recording studios. MIDI, an acronym that stands for **Musical Instrument Digital Interface**, allows us to connect devices together that control sound (keyboard controllers and control surfaces) and make sound (synthesizers, samplers, and virtual (software) instruments). The computer digital audio workstation (DAW) is “inserted” between the sound controllers and sound producers. The DAW records digital instructions for creating sound to instrument tracks. During playback of these tracks, these digital instructions are sent to the devices that produce the desired sound. Quite often, the sound producing devices are virtual instrument plug-ins within the DAW itself.

Shown in the title figure above, a keyboard controller is often the source of the MIDI digital instructions that are recorded to instrument tracks. In my case, my keyboard controller is my Yamaha Clavinova digital piano. I connect a 5-pin MIDI cable from the MIDI output of my Yamaha keyboard to the MIDI input of my **PreSonus Studio 68c audio interface unit**. The PreSonus audio interface unit connects via USB-C to my computer running the **PreSonus Studio One 6 DAW**.



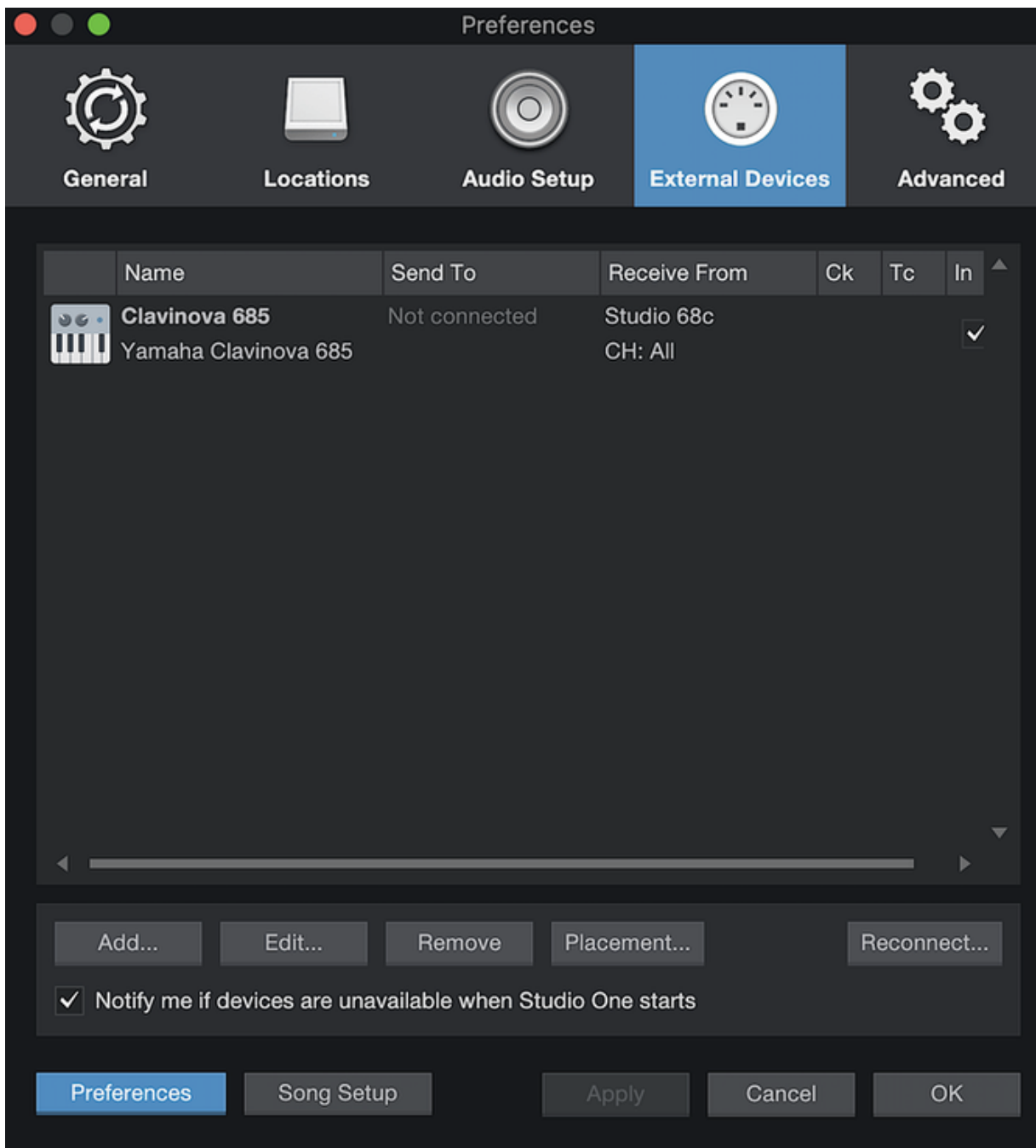
MIDI Output from Yamaha Clavinova Digital Piano



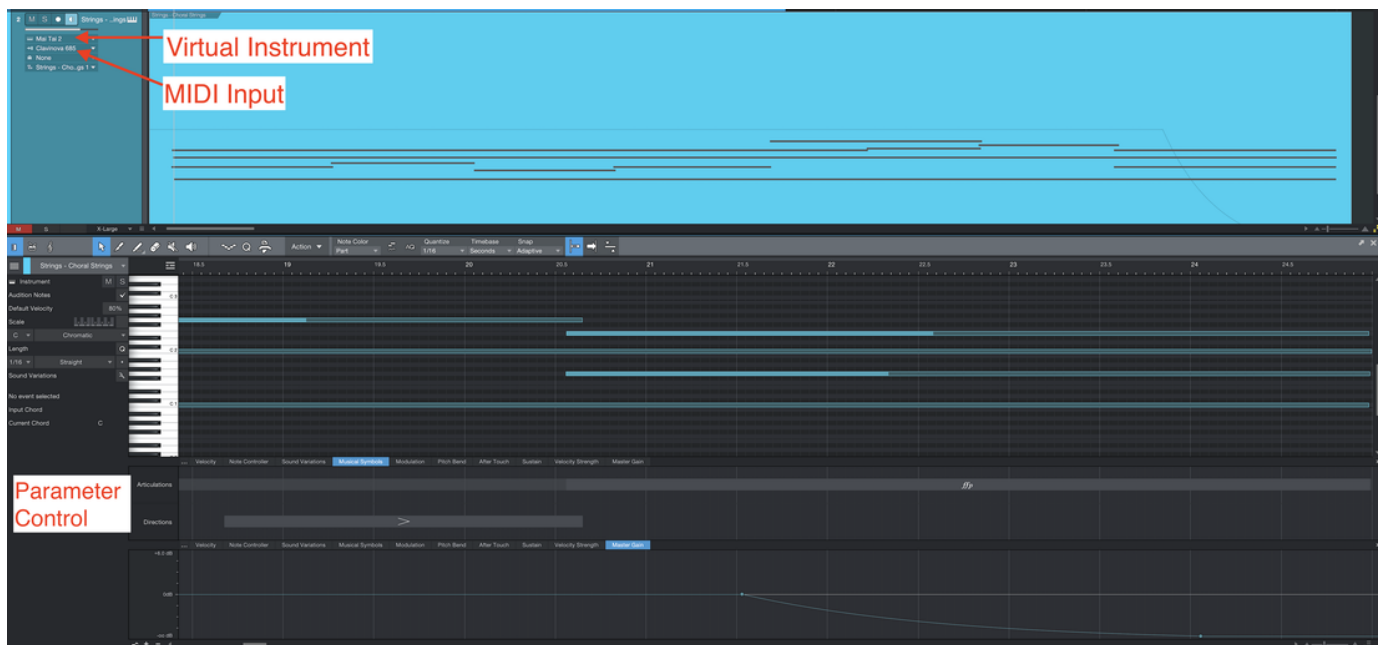
MIDI Input to PreSonus Studio 68c Audio Interface

The next step is to configure the PreSonus Studio One DAW to accept the MIDI keyboard controller instructions as input to an instrument track. Navigate to *Preferences/External Devices/Add Device*, and configure the following items:

External Device Name: Yamaha Clavinova 685
Receive from: Studio 68c (audio interface)
Channel: All
Check box as Default Instrument Input
Enable MIDI Polyphonic Expression (MPE)



Now we're ready to record MIDI data to an instrument track ! In the figure below, the recorded note data are shown in the instrument track, and the accompanying control parameters (such as note velocity, articulation, and after touch) are shown in the parameter automation lanes in the lower half of the window. The MIDI input channel is the Clavinova 685. The virtual instrument plug-in is the Mai Tai 2 synthesizer with Choral Strings sound settings. This recording is a good example of putting down a synth pad track.

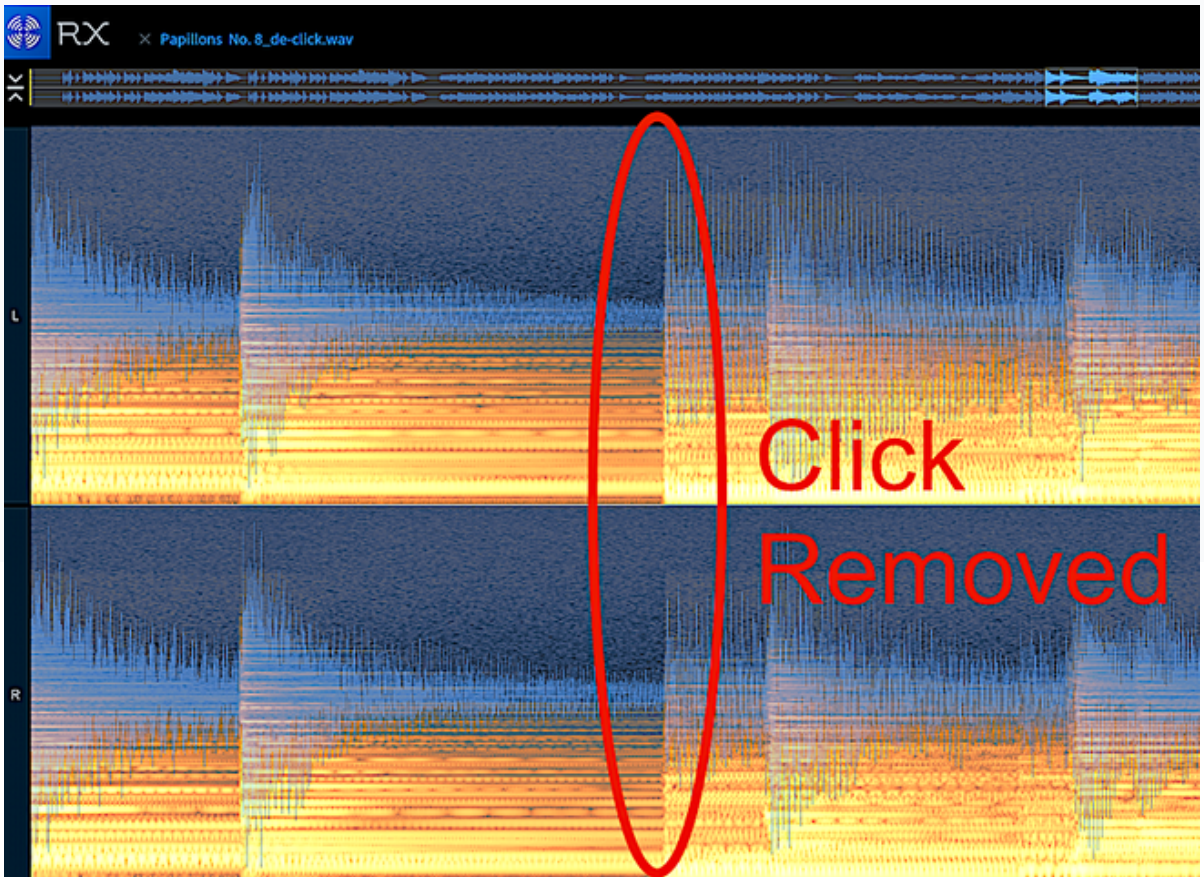


[Click on image to enlarge](#)

A Final Word :

Since multi-tracking of MIDI instruments is one of the most popular activities done in home music recording studios, we need to remain mindful of the issues of latency and driver error compensation affecting the overdubbing process during recording. In the previous chapter, calibrating your DAW's **driver error compensation** using a loopback test and record offset for MIDI instrument tracks was discussed.

B11. Audio Repair



It is quite common to have various sound artifacts present in your recordings from a variety of sources, with the effect of these artifacts on the sound quality of the recording ranging from mildly annoying to absolutely devastating. Here are some of the more common kinds of sound artifacts:

- Broadband noise
- Clicks and pops
- Low-frequency hum
- Distortion from clipping
- Vocal plosives and sibilance

The industry-leading audio repair software, **iZotope RX 10**, can literally “fix” these and other problems. At the heart of this software application stands a real-time spectrum analyzer coupled with intelligent repair algorithms powered by machine learning. In my home studio, I have the budget-friendly iZotope RX Elements version of this incredibly powerful tool. This RX Elements version is a stripped down (lite) version of the more comprehensive RX 10 Standard and Advanced versions, and includes a stand-alone editor application with four essential repair modules :

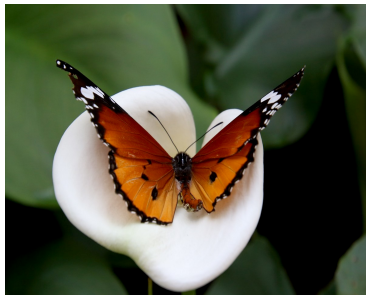
- De-Click
- De-Clip
- De-Hum
- Voice De-Noise

The application also contains a number of audio waveform editing tools that include fade, gain, stereo and phase controls. There is a learning curve in figuring out how to use this audio repair editor -- I leave that task to the large number of tutorial videos available online. For my purpose here, I'd like to show an example of removing a couple of annoying “clicks” that are present in the audio file of my recording of Robert Schumann's Papillons No. 8 piano piece. These two clicks occurred at two points in time when sound events were edited with inadequate cross-fading between them. So, I'd like to fix my errors using post-production audio repair with the iZotope RX De-Click module.

Shown below are the original sound file with the clicks and the repaired sound file with the clicks removed. You can see how these transient clicks are visually represented in the spectrogram as very broad frequency bandwidth energy impulses. These two clicks occur at about the 7 second and 17 second marks on the time axis. In the repaired file, notice the absence of these broadband impulses in the spectrogram ! Preferably on earphones, listen carefully to the two sound clips and see if you can notice the difference.

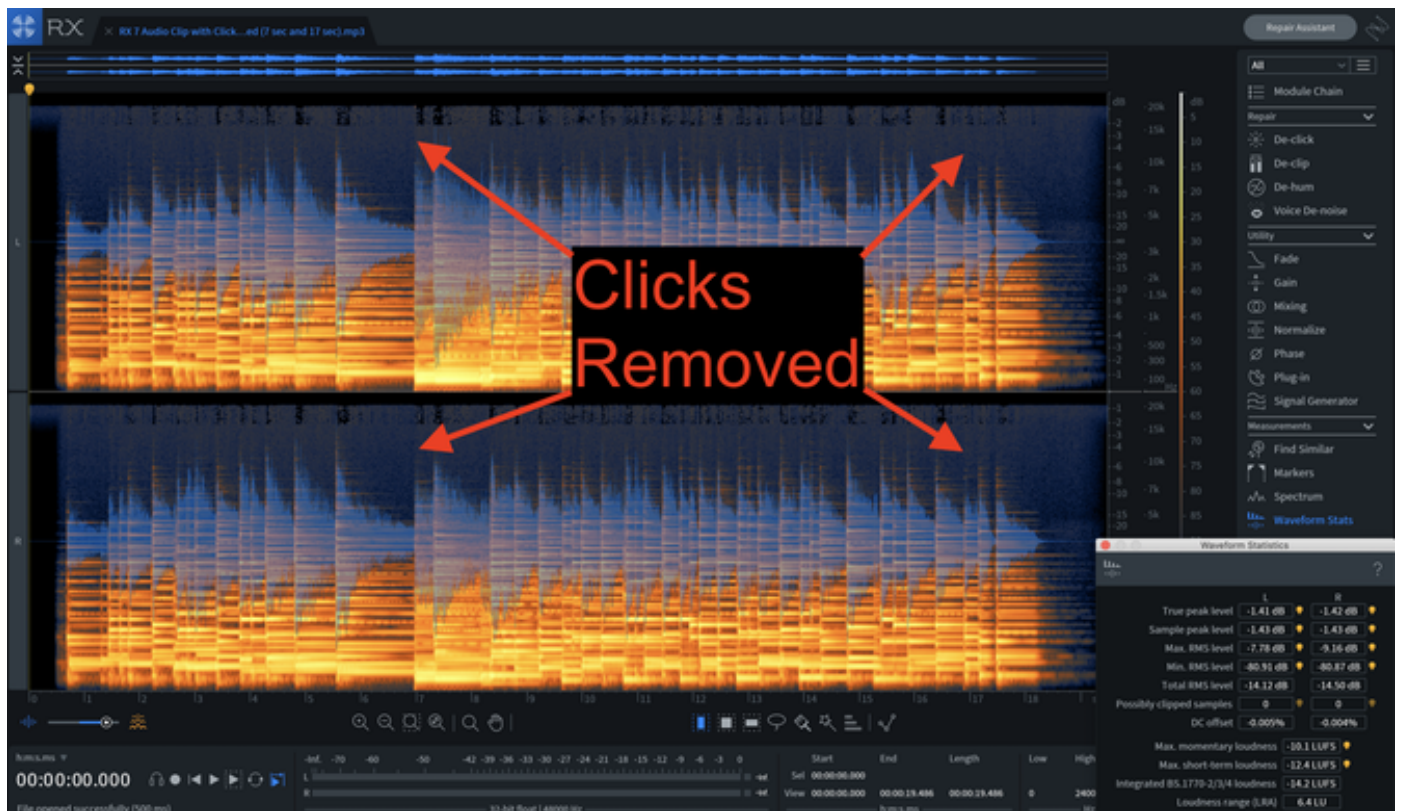


(Click on image to enlarge)

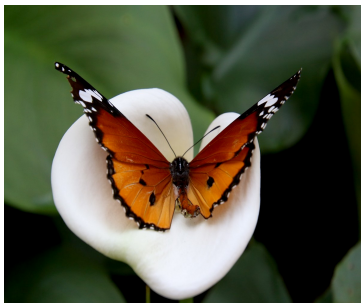


Audio File with clicks at
7 seconds and 17 seconds

Schumann Papillons No. 8 Piano Piece



(Click on image to enlarge)



Audio File with clicks **removed**
at 7 seconds and 17 seconds

Schumann Papillons No. 8 Piano Piece

The iZotope RX Elements software came bundled (free) with other iZotope products that I was interested in using. At the time, I ignored the RX Elements application, thinking I would never really need to use it. But now that I have had time to play around with it, I can see just how powerful and useful these audio repair modules are !! So, it just goes to show that there is always something new to learn that will help you produce higher quality sound recordings in your home music studio.

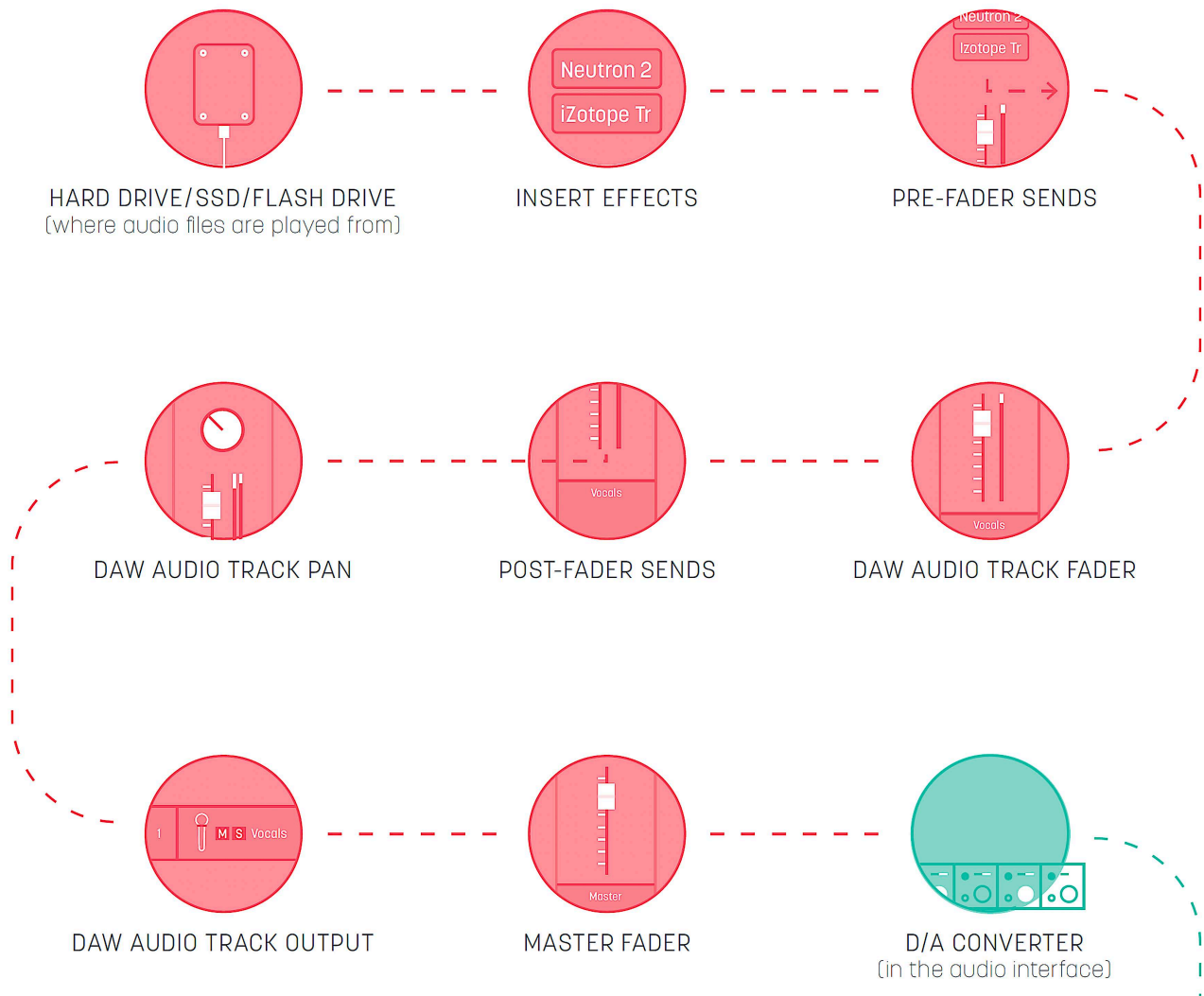
C. Mixing

C1. DAW - Mixing



[Click image to enlarge](#)

In Section C. Mixing, we'll take a look at the various kinds of signal processing done in the **digital audio workstation (DAW)** to produce a final “mix-down” of the music. Simple block diagrams of the signal flow in a DAW are available for download from **iZotope**. The signal flow chart for mixing in the DAW is shown below.



The signal processing blocks following the solid-state drive (SSD) are concerned with the monitoring/playback process, and have no effect on the recorded audio files. For example, the DAW audio track fader does not affect the amplitude level of the recorded signal. The signal flow in the chart above depicts the flow down the channel strip of a hardware mixing console or the DAW mixing panel. Playback of a track originates from the recording on the SSD. The signal then passes through these basic elements of signal processing and mixing:

Insert Effects: Dynamics and Equalization Processing

Audio Track Fader: Mix Level Control

Pre- and Post-Fader Sends: Time-Based Effects

Audio Track Pan: Stereo Image Processing

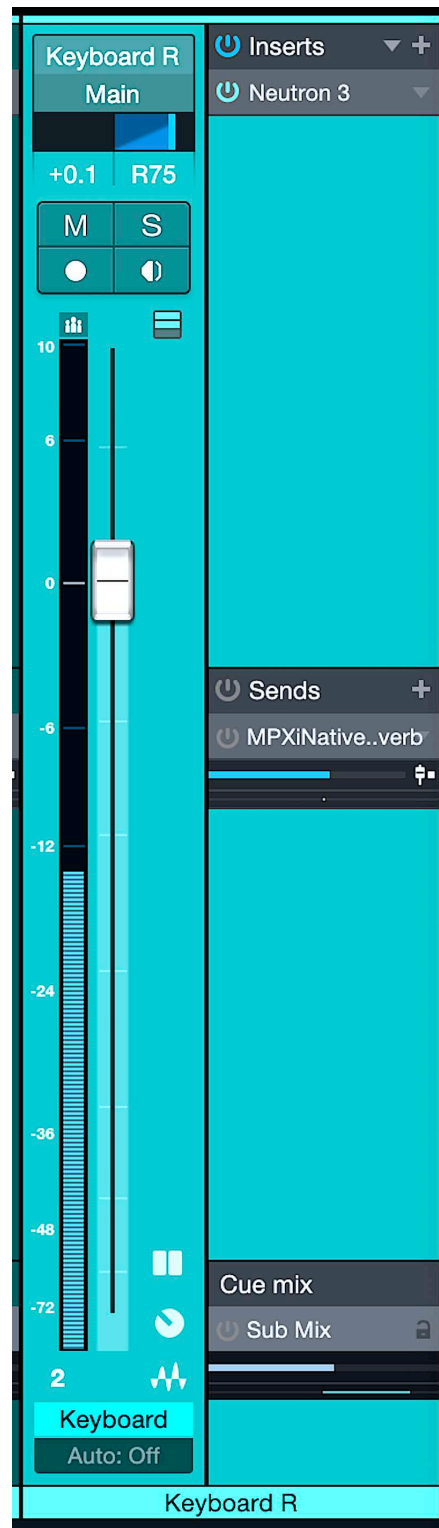
Main Output Bus: Stereo Mix

Master Fader: Main Output Level Control

Main Output Bus: Mastering Effects

The final mixed digital stereo signal either leaves the computer and returns to the audio interface unit for playback, or is “bounced” (written) to a formatted audio file, such as a WAV, FLAC, M4A or MP3 file.

One channel strip of the **PreSonus Studio One mixer console** is shown below.



Channel Strip

As described in the chapter on **recording in the DAW**, the input section sends the digital signal waveforms to the SSD, where each waveform is recorded as an individual track. In the example shown above, the “Keyboard R” (right stereo channel of the keyboard) track output from the SSD becomes the input to the DAW channel strip, where the signal processing and mixing will be done. The channel “Inserts” are the first signal processing elements to occur in the signal path chain.

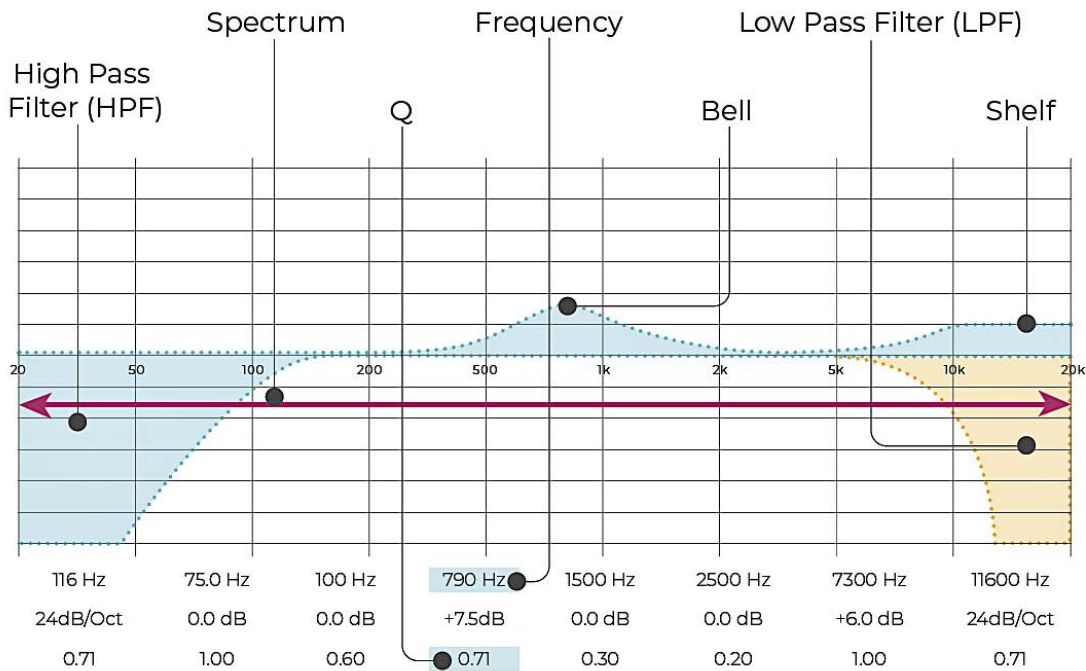
Equalization (EQ) and Compression Inserts are the subjects of the next two chapters.

C2. Equalization



The “mixing” of recorded tracks is done using the mixing console of the Digital Audio Workstation (DAW). Each recorded track is brought into the mixing console as the input to its own channel strip, where the signal processing and mixing occurs. As shown in the previous chapter, the channel “Inserts” or “plug-ins” are the first signal processing elements in the signal path chain. Typically, the two most common plug-ins are Equalization and Compression. Equalization (EQ) is the subject of this chapter.

EQ shapes the tonal balance and overall sound quality of the recorded music, and is perhaps the most important artistic part of the entire mixing process ! Applying EQ properly requires great skill and much practice (good ears) for the sound engineer. There is more on-line and studio class instruction and advice on EQ than on any other topic in sound recording. I have found a good place to start learning about EQ can be found on the website **Musician on a Mission (MoaM)**. MoaM also offers excellent YouTube videos, Masterclasses, and free “cheat sheets” on EQ and mixing.



[Click on figure to go to MoaM EQ tutorial](#)

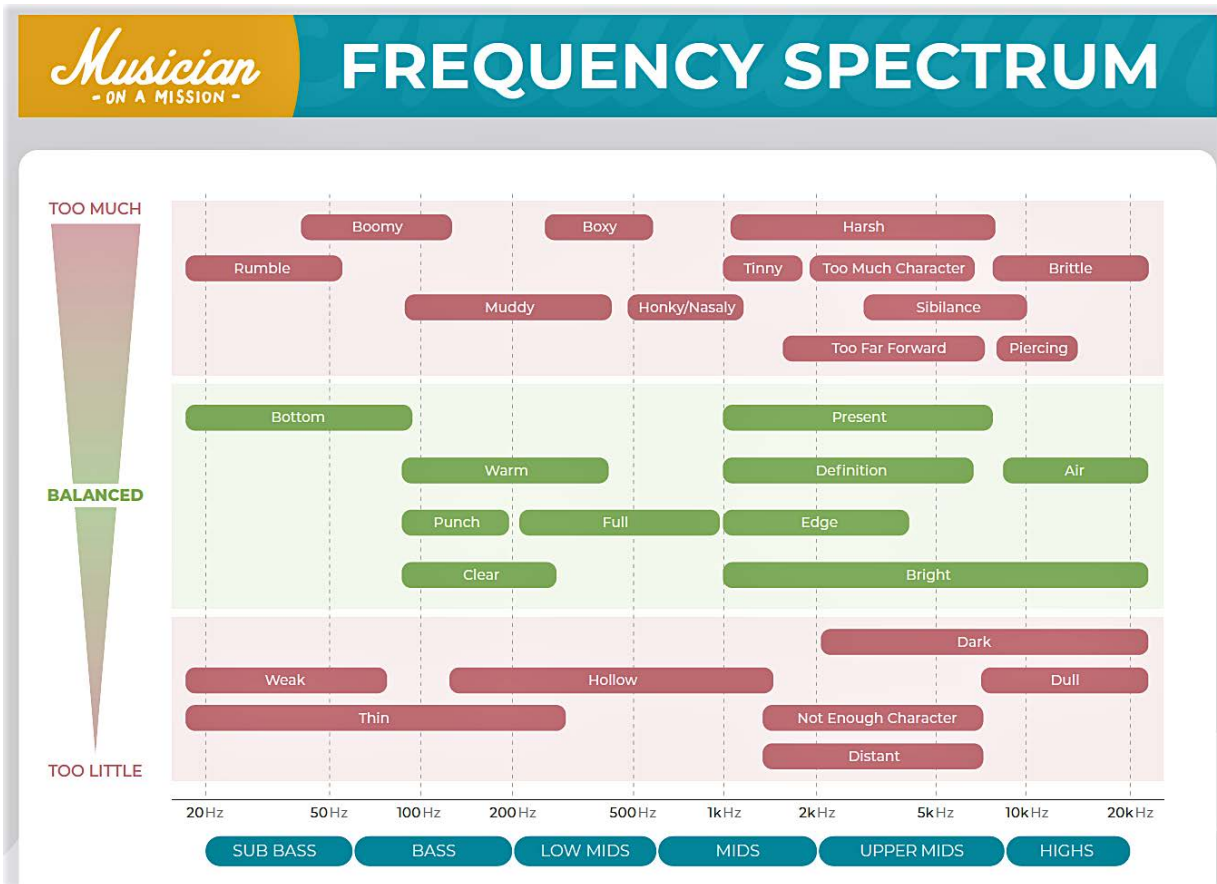
An EQ is a set of electric filters that cut or boost the signal spectral amplitude around a range of frequencies. Typically, these filters include high-pass, low-pass, bandpass (bell), and shelf filters, as indicated in the spectrum graph above. In very simple terms, there are three general strategies for employing these filters:

1. Removing unwanted spectral energy.

Using narrow-band bell filters ($Q\text{-factor} > 2$), you can surgically cut out artifacts, such as room resonances. Also, high-pass filters can be used to remove low-frequency noise and bass rumble and boom. By eliminating these, you “make room” in the spectrum for the desired signal tones.

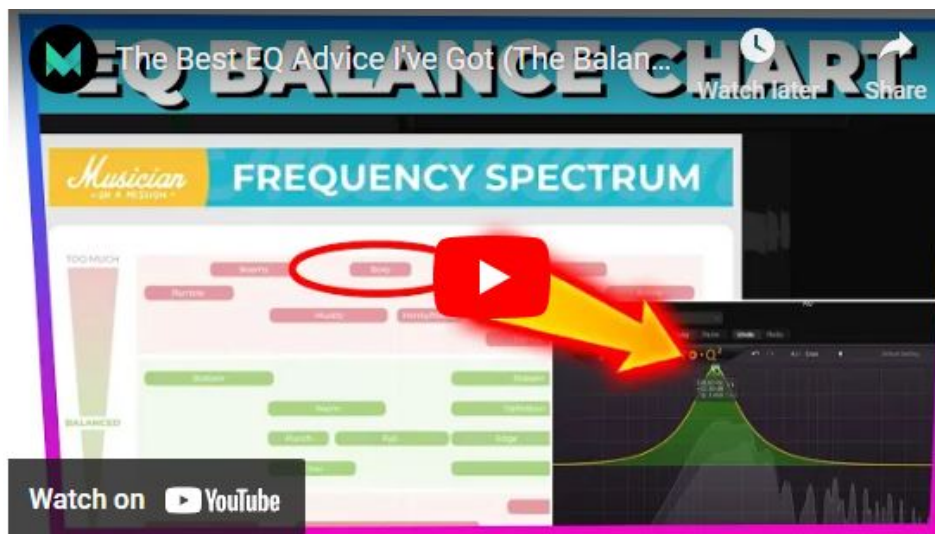
2. Boosting/cutting spectral energy in the different frequency regions.

Using moderate-bandwidth bell filters ($Q\text{-factor} < 2$), you can enhance or lessen the spectral amplitude in a frequency band to achieve the desired sound quality and tonal balance. Click on this link to get an extremely useful **EQ Balance Chart** from MoaM.



EQ Balance Chart

By using adjectives to describe the sound that you're hearing, you can adjust (boost or cut) the energy in a band to achieve a good tonal balance and sound quality. As an example, adjusting the amount of energy in the bass / low-mid ranges (150 – 450 Hz) can yield a sound that is clear and full, and not too thin or too muddy. Here's some really awesome EQ advice on using the EQ Balance Chart from the folks at MoaM - definitely the best video I've ever seen on this topic !

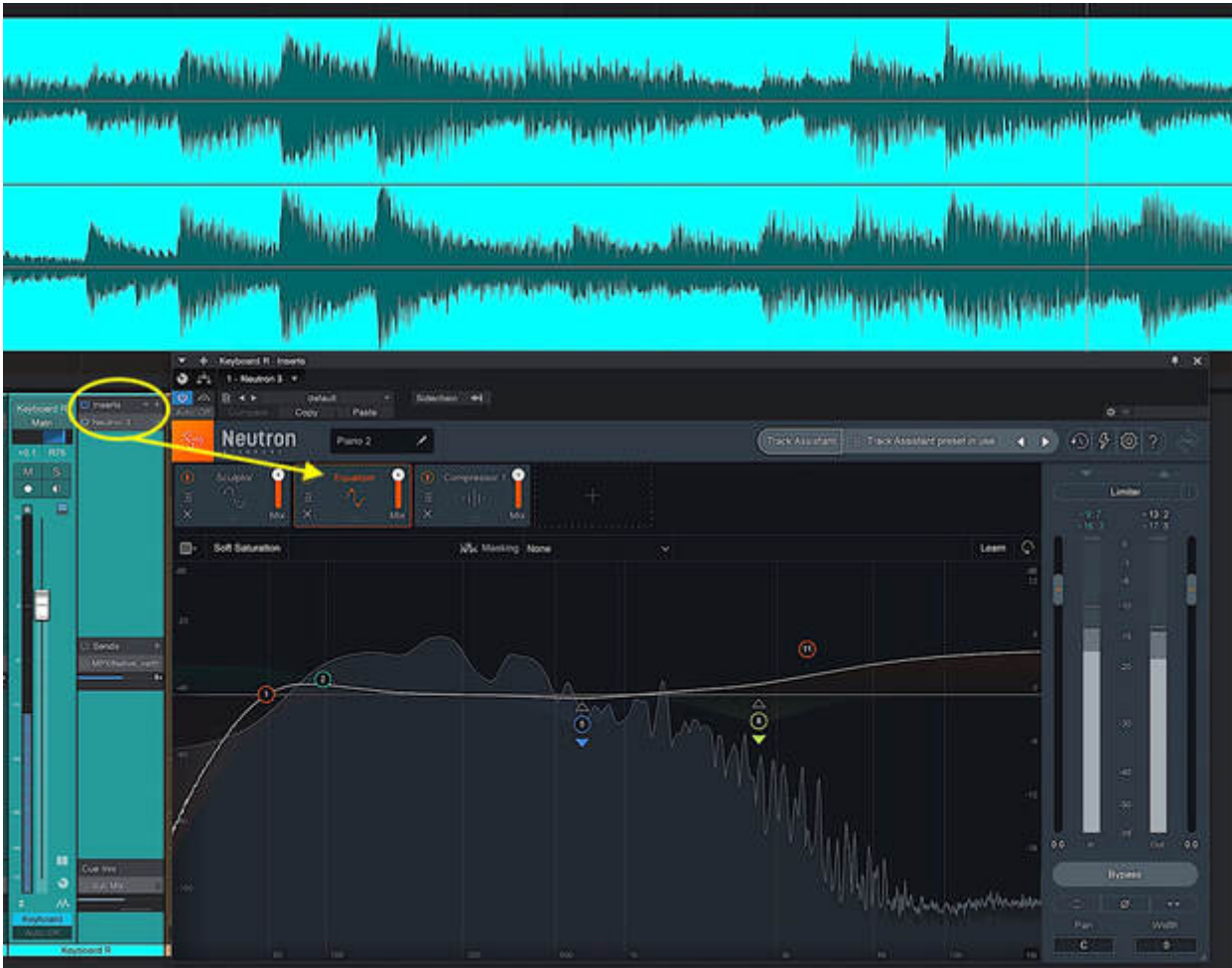


3. Creating space in the mix.

When two or more instruments have their primary spectral energy in the same frequency band, there can arise the problem of “masking”. As its name suggests, masking can hide a more “important” instrument under other instruments. So, you want to create separation and clarity in your mix by “allocating” a frequency range to that important instrument. By cutting frequencies in some instruments, and boosting them in others, you can create space in the mix and give each part its own place to reside in the frequency spectrum.

These are the three basic strategies in the EQ process. There is, however, a host of other EQ skills and techniques that need to be learned and utilized. For example, it is better to apply EQ to a track while listening to the changes in the context of the whole mix, *i.e.*, avoid applying EQ to a track in solo. Another concept to be aware of is the “see-saw” effect. A cut/boost in a narrow frequency range can be accompanied by a boost/cut in a narrow frequency range on the other side of the spectrum – hence the “see-saw” effect ! For example, to reduce muddiness, you may try cutting the lower-mid frequency range, but it doesn’t quite seem to alleviate the problem enough. On such an occasion, you can also try boosting the other side of the spectrum – the higher-mid frequency range in this example.

As I mentioned before, EQ sits at the heart of mixing. It takes time and plenty of experimentation and practice to develop a “good ear” capable of achieving just the right tonal balance. I try to use a minimum of EQ in my mixes of classical piano pieces – mostly out of concern that too much EQ will degrade the already wonderful sound quality of the Yamaha digital grand piano ! The **Izotope Neutron EQ plug-in** is shown here in the Insert of the Keyboard R channel strip.



Izotope Neutron EQ in "Over the Sea to Skye" track (Click to enlarge)

In the EQ window, the real-time spectral amplitude of the piano is displayed against the EQ envelope formed by the 5 filters: bass roll-off high-pass filter (Band 1) and bass bump-up shelf filter (Band 2), dynamic cut bell filters (Band 5) and (Band 8), and high-frequency bump-up shelf filter (Band 11). The dynamic bell filters have three EQ parameters you're used to seeing – center frequency, Q (bandwidth), and gain – with the addition of a threshold parameter like you're used to seeing on compressors. When spectral amplitude goes above (or below) this threshold in the specific band, it triggers a cut (or boost) in the gain of the bell filter. This combination allows you to use EQ in a way that responds and adapts to incoming audio, only affecting it when the amplitude in that frequency band crosses the threshold. While an instrument performance might be pretty smooth overall, there could be moments when a certain frequency (resonance) pokes out of the mix in a distracting way. Since that frequency only needs to be attenuated when it gets too loud, we can use dynamic EQ to tame this in real time.

In the next chapter, we'll take a look at the next Insert in the signal path – the Compressor plug-in.

C3. Compression

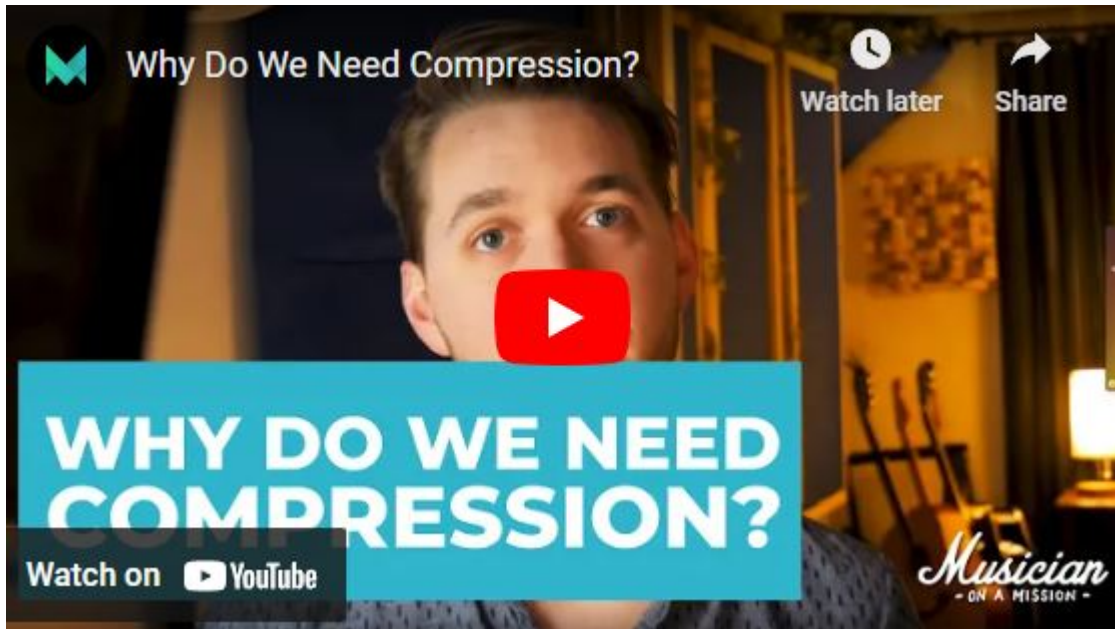


[Click on image to enlarge](#)

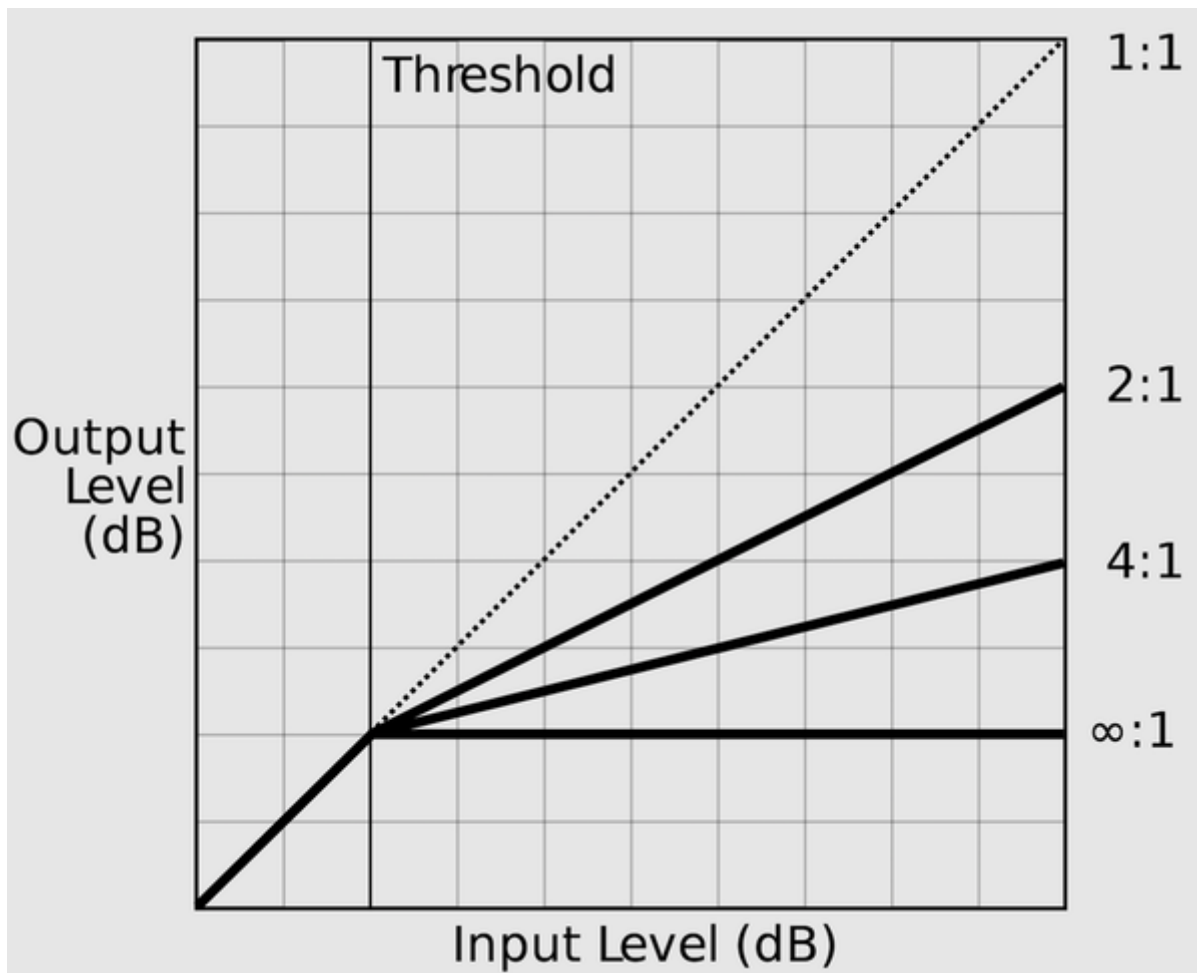
In the previous chapter, we looked at **Equalization** to achieve tonal balance in the sound mixing process. In this chapter, we'll take a brief look at **Compression**, which works to balance dynamics in the mix.

In the figure above, you can see the Compressor plug-in following the EQ plug-in for my **iZotope Neutron Insert**. There has been much said about the order of these plug-ins in the signal path – EQ or Compressor first? Some good advice on this is provided in a video by **Joe Gilder of PreSonus**. My "take-away" is that it is usually better to EQ first, so that you take care of tonal imbalances, resonances, and masking before you compress. Compression, if needed at all, then gives your track a smooth, well-balanced dynamic sound.

A really good overview of four reasons for using compression can be found in the video below from the folks at **Musician on a Mission**.



The primary function of compression is to reduce the high amplitude (loud) portions of the audio waveform. When a set input Threshold level is exceeded, the output level no longer 'tracks' the input level – it is reduced by a fraction. For example, with the Ratio parameter set at 4:1 on the Compressor plug-in, when the input level rises 4 dB above the threshold level, the output level will rise only 1 dB .



Output Level vs. Input Level in a Compressor Plug-In

In order to compensate for the reduced sound volume, make-up gain is applied to the entire track – consequently also bringing up the volume of the quiet portions of the sound waveform. In essence, the dynamic range (maximum-to-minimum amplitude ratio) of the sound has been reduced, or compressed. This makes the volume of the track more consistent and makes the musical performance sound smoother.

It may seem that a reduction in dynamic range is a bad thing, since dynamic changes are a large part of the excitement and energy in music. In certain musical genres, like rock, sacrificing excessive dynamic range in the numerous and separately recorded instrument tracks is needed to make a good balance in dynamics in the overall mix. In professionally mixed rock songs, compression is almost always used on the drum, guitar, and vocal tracks. Dynamic changes through the song (between chorus and verses, for example) can be adjusted using volume automation of the faders.

In a musical genre like Classical music, little, if any, compression is generally needed in the mixing phase. There is usually only a single stereo track and a few mono tracks in the recording of a solo instrument, a chamber group or an entire orchestra. The dynamic balance between instruments is accomplished by the skilled performance of musicians perfectly blending the sound, and the full dynamic range of the recorded music should be preserved.



PreSonus Compressor in the Studio One DAW

Shown in the figure above is one of the **PreSonus Compressor software plug-ins** that is included in their Studio One DAW. The main parameters to control when using any Compressor plug-in on a track are:

Threshold:

This is the input signal amplitude level above which the compressor becomes engaged.

Knee:

At the threshold, there is a discontinuity in the slope of the input level / output level curve (see Compression graph above.). The “knee” parameter smooths out this discontinuity, so that there is a gradual transition between compressed and un-compressed audio.

Ratio:

This sets the amount of compression, from light compression (2:1) to heavy compression (20:1).

Make-up Gain:

This is gain applied to the signal to “make up” for the decrease in sound volume caused by compression. Since this also increases the sound volume of quieter portions of the waveform, the overall effect of compression is to reduce the dynamic range of the audio.

Attack Time:

This is the time delay (in milliseconds) from when the signal level exceeds the threshold to the activation of the compressor. This allows for initial transients in the signal to escape being compressed, leading to a sound with more “punch” and presence.

Release Time:

This is the time delay (in milliseconds) from when the signal level falls below the threshold to the de-activation of the compressor. This helps avoid ‘pumping’, the audible unnatural level changes associated primarily with the release of the compressor.

We’ve now looked at achieving balance in tonal character and dynamics in the mix. Next up, we’ll take a peek at spatial balance in the mix, using time-based effects.

C4. Reverberation



Royal Birmingham Conservatoire Concert Hall

In the previous two chapters on **Equalization (EQ)** and **Compression**, we talked about achieving balance in tone and dynamics in the mixing process. In this chapter, we'll take a brief look at a time-based effect, **Reverberation**, that works to bring spatial balance to the mix.

We listen to music in space – acoustic waves reaching our ears from the immediate environment around us. The spatial aspect of sound can be simply characterized by its width and depth. The width is primarily associated with the stereo image – Left/Right or Mid/Side – and is balanced using panning on individual tracks and using stereo image enhancing plug-ins. The depth of the sound is mostly associated with the time-domain delay of the sound reaching our ears. Depth gives us a feeling of the size and character of the space, e.g., a stone cathedral versus a small living room.

Sound waves reach us by three “paths” :

- 1. Direct.** This is the direct ‘line of sight’ from source to receiver.
- 2. Low-order Reflection.** This occurs when sound waves from the source undergo one or two reflections before reaching the receiver.
- 3. Multiple Reflections.** This occurs when sound waves from the source undergo many reflections before reaching the receiver.

Direct waves arrive at the receiver first in time. Reflected waves arrive at the receiver at delayed times, since they travel different distances. The nature of the reflections plays a key role in determining how long the sound will reverberate in the space. For example, reflections from a smooth flat hard surface will be very different from reflections from an irregularly-shaped absorptive surface. In the field of acoustics, the Sabine equation gives an approximate “reverberation time” (time for the sound level to decay by 60 dB). This simple equation depends on the volume of the space, the surface area of the enclosure, and the various wall material absorption coefficients.

Sabine’s Formula

Sabine’s formula is given by the following:

$$RT_{60} = \frac{24 (\ln 10) V}{c_{20} S_a}$$

RT_{60} is the reverberation time (to drop 60 dB)

V is the volume of the room

c_{20} is the speed of sound at 20°C (room temperature)

S_a is the total absorption in sabins

The sabin unit has the same dimension as area (e.g. m²). A one square meter surface with an absorption coefficient of 0.75 would be considered 0.75 sabins. The absorption coefficient has a range of 0 to 1, where a coefficient of 0 indicates none of the sound is absorbed, and a coefficient of 1 indicates that 100% of it is absorbed.

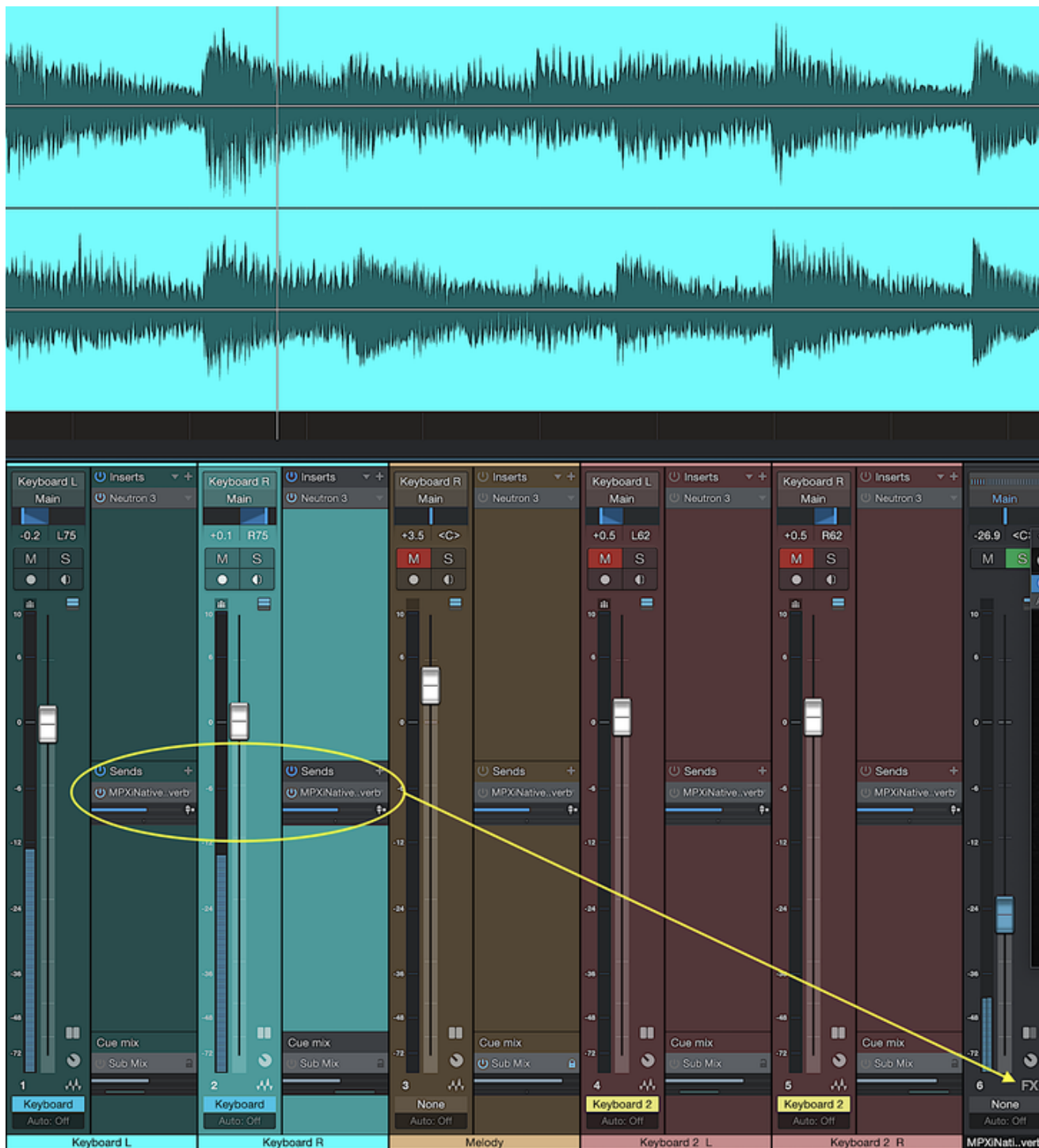
Since we know the speed of sound at 20°C is 343 m/s, we can do a little math and reduce the formula to:

$$RT_{60} \cong \frac{0.161\text{s/m } V}{S_a}$$

For great music concert halls, such as Carnegie Hall and Boston Symphony Hall, reverberation time RT is on the order of 1.5 – 2.0 seconds.

Adding Reverb to your mix puts listeners in a certain space, controlling where they are in relation to the music. Reverb is often described as adding warmth or 'wetness' to the sound. Less reverb on a track brings out the direct sound, bringing the instrument 'up front' and making it clear and present. More Reverb on a track increases the depth of the sound, pushing the instrument 'to the rear' and making it distant and dreamy. Reverb should be used sparingly, since too much reverb 'muddies' the tonal balance. Reverb plug-ins often have built-in EQ that reduces the low-mid frequencies to lessen this muddiness.

In your mixer console, Reverb is applied to the music by creating a bus FX (Reverb) channel and using AUX Sends from your instrument channels to this bus FX channel. Typically, these AUX Sends are post-fader sends, *i.e.*, the signal sent to the FX bus is *after* the fader adjustment of the channel volume level. This assures that the "Wet/Dry" mix of the signal remains at a fixed ratio for any changes in the channel fader position. An example of AUX Sends to the FX bus in the **PreSonus Studio One DAW mixer** is shown below. The output of the Reverb channel goes to the Main output bus and is mixed with the instrument channels there.



Instrument channel AUX Sends to FX (Reverb) bus. (click image to enlarge)

There are many different Reverb plug-ins available from a host of audio FX developers. The digital signal processing performed in the Reverb software is based on either algorithmic or convolution computations. The algorithmic approach uses mathematical 'formulas' to synthesize delayed waves that are similar to those that occur in acoustic environments. An algorithmic Reverb plug-in requires only moderate CPU processing power.

The convolution approach uses actual sampled impulse responses of real acoustic environments to calculate the response of the system (the acoustic environment) to the driving source waves. (This mathematical calculation is based on the convolution integral of Green's functions.) The big advantage of a convolution Reverb plug-in is that it accurately produces the natural sound of real acoustic environments. The drawback, however, is a huge computational burden on CPU processing power.

The Reverb plug-in that I use is a very simple one – the classic **Lexicon MPX Native Reverb** (an algorithmic reverb).



Lexicon MPXi Native Reverb plug-in (click image to enlarge)

There are one-hundred Reverb presets to choose from – this allows you to “dial in” a descriptive spatial quality without having to adjust all the individual parameters. Examples of presets include “midnight chamber” (in figure above), “stone cathedral”, and “large living room” .

The main parameters to adjust when using any Reverb plug-in on a track are:

Reverb Type: different sizes of Hall, Chamber, Room, Stage, Theatre, Church, or Cathedral.

Pre-delay: time between the arrival of the direct sound from the source and the arrival of the first sound reflection. (milliseconds)

Reverb Time: time for the amplitude of the sound reflections to decay by 60 dB. (seconds)

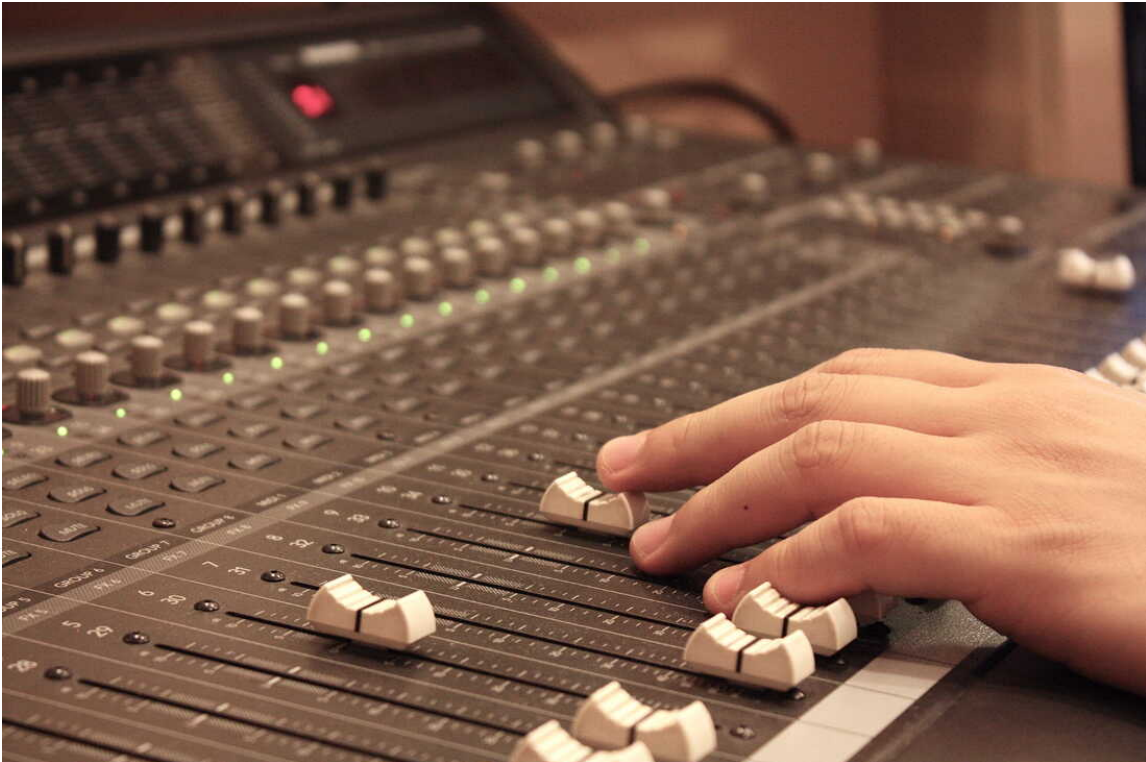
Diffusion: amount of dispersion of the sound waves due to complex surface shapes in the space – leads to many reflections arriving at the receiver over a full 2π steradian solid angle. Increasing diffusion has a smoothing and thickening effect on the reverberation. Reducing diffusion tends to separate discrete reflections in time, giving a more open sound (echoes).

Wet/Dry Mix: percentage mix of the “wet” signal (with reverb) to the “dry” signal (without reverb). Since we are using an FX bus channel to send the reverb (wet) signal to the Main output bus where it mixes with the original (dry) signal, we adjust the wet/dry mix using the Aux Send level control and FX bus channel fader. Therefore, we can set the wet/dry mix parameter of the Reverb plug-in unit to 100% .

There are other time-delay effects based on replications of the signal that can be applied, such as delay, flanging, doubling, and chorus. Reverb, in my opinion, is the most “natural” effect, as it mimics the acoustics of real listening environments. Careful and intentional application of Reverb to your mix will greatly enhance the presence and liveliness of your music.

In the next chapter, we’ll continue along the signal processing path in the mixer console channel strips to the panning and volume fader controls.

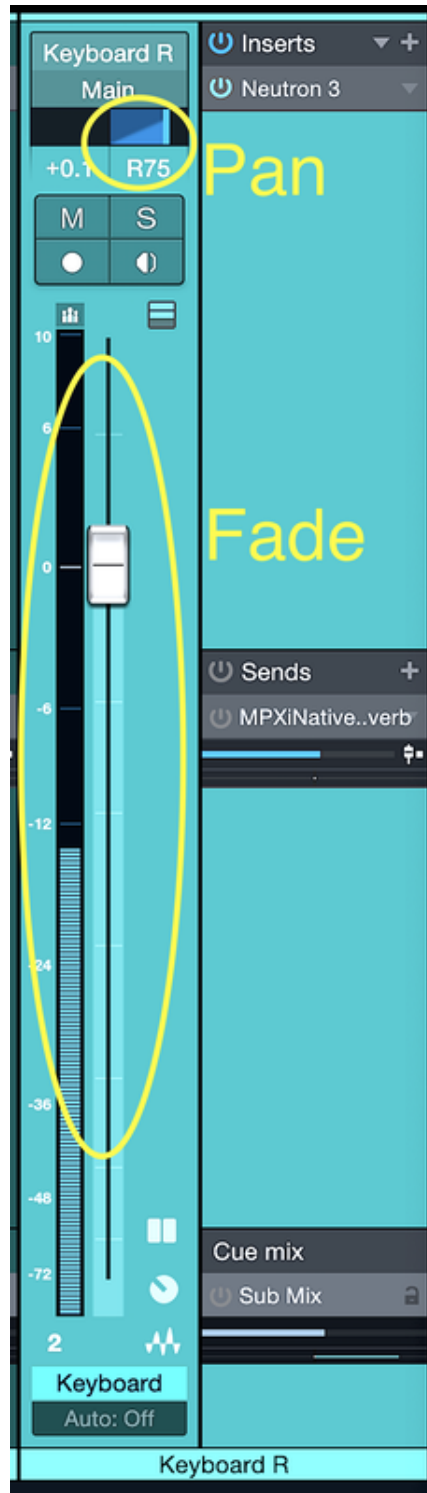
C5. Fade and Pan



The final signal processing done in each mixer channel is fading and panning. Fading sets the output volume of the track to the main mixer bus, and is the key process in properly staging and blending (mixing) the individual music tracks together in a sound recording. Panning sets the Left-Right (or Mid-Side) image of the track in the stereo main output.

Fading

First, we should note an important difference in the terms “gain” and “volume”. Typically, **gain** refers to the input sound level (dBFS) to the mixer channel, whereas **volume** refers to the output sound level from the channel. Gain was discussed in some detail in two prior chapters on adjusting track gain and gain staging -- **Digital Signal Levels** and **DAW – Recording**. As mentioned above, each channel strip contains an associated channel fader that adjusts the signal output level sent to the main mixer bus. The fader control in one channel strip of the mixer panel of the **PreSonus Studio One DAW** is shown below.



Typically, the fader level is initially at 0 dB (unity gain), and adjustments are made to this level to stage and blend this instrument track with other instrument tracks that are summed together on the main bus. Such adjustments are usually to attenuate the signal level (negative dB values), but small increases in signal level ($< +3\text{dB}$) can be applied sparingly.

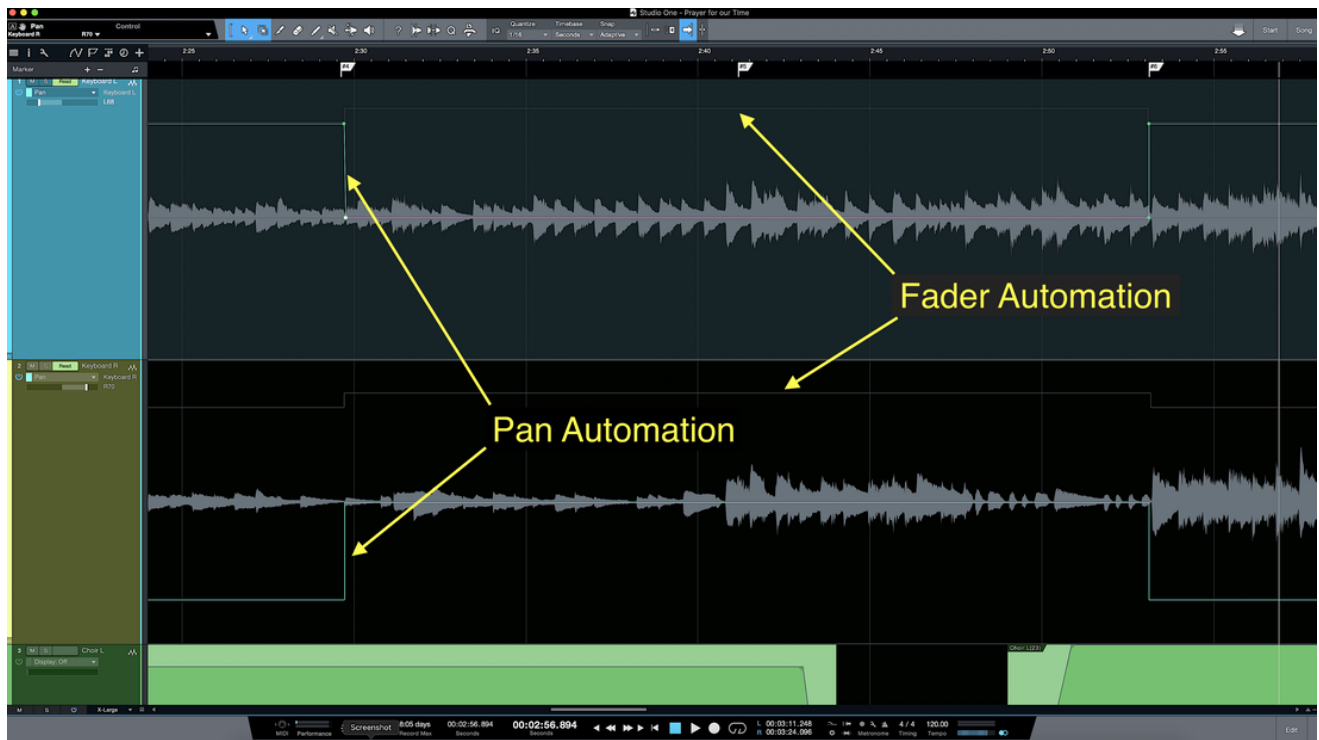
In recordings with many instruments and/or many input channels, we can use “grouping” to make a **submix** . A number of channels can be organized into a group, such as all the microphone inputs from the drum kit (which can be quite a lot!). Such grouping allows the relative levels of the individual channels to be interlinked. This feature makes it possible for the multiple channels to retain their relative level balance (the submix) while offering control over their overall group level from a single fader or stereo fader pair.

Panning

Panning sets the spatial positioning of the sound in the stereo image. This is left-right positioning, or sometimes designed to be mid-side positioning. The width of the spatial sound is determined by the panning of all the channels in the mix. During mastering, a process that follows mixing, the width of the full stereo image will be analyzed in terms of correlation of sound waves at the listener, and slight adjustments to this width may be made at that time.

Automation

The process so far creates a static mix, *i.e.*, the relative fader positions and panning pot settings remain fixed throughout the musical recording. However, both fading and panning can be **automated** so that these settings can change in time to follow the needs of the music. In a digital audio workstation (DAW) mixer, automation of faders and panners is accomplished simply by drawing “rubberband” envelopes on the screen alongside the sound waveforms in each track. An example is shown here for two tracks in the **PreSonus Studio One DAW**.



Automation of fader and pan controls (click image to enlarge)

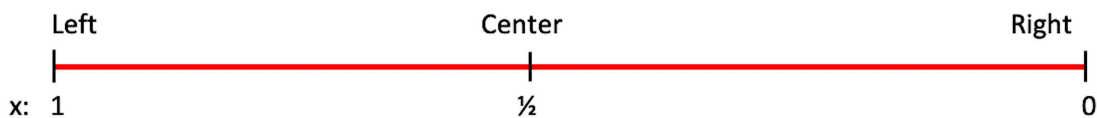
Finally, I should mention that most signal processing effects, such as EQ, Compression and Reverb, can also be easily automated in your DAW, thereby making these effects dynamic in real time.

Over these last five chapters, a brief overview of the signal processing done during the sound mixing phase of making a musical recording has been presented. In the next few chapters, I'll discuss the Mid/Side and imaging processes for setting up the stereo field in stereo tracks.

C6. Mid/Side Signal Processing



A Mid/Side signal processing plug-in may be used on a stereo channel to widen (and sometimes narrow) the stereo sound image. How this imaging plug-in works is the topic of this chapter. Mid/Side signal processing is a clever technique that is used for stereo microphone recordings, mixing tracks, and mastering of the main output bus. To get a sense of the basic concept behind Mid/Side processing, I've used some simple algebra to track the signal voltage amplitudes of the left and right stereo channels through the imaging plug-in:



Fraction of signal panned Left: x $0 \leq x \leq 1$
Fraction of signal panned Right: $1-x$

Left (L) Channel: $L = x$
Right (R) Channel: $R = 1-x$

The three steps to Mid/Side processing in the imager plug-in are:

1. Transform from Left (L) and Right (R) channels to Mid (M) and Side (S) channels:

Mid channel is Sum of the Left and Right channels $M = L + R = x + (1-x) = 1$

Side channel is Difference of the Left and Right channels $S = L - R = x - (1-x) = 2x - 1$

Examples:

Center Pan:

$$\begin{aligned}x &= \frac{1}{2} \\L &= \frac{1}{2} \quad R = \frac{1}{2} \\M &= 1 \quad S = 0\end{aligned}$$

Hard-Left Pan:

$$\begin{aligned}x &= 1 \\L &= 1 \quad R = 0 \\M &= 1 \quad S = 1\end{aligned}$$

Hard-Right Pan:

$$\begin{aligned}x &= 0 \\L &= 0 \quad R = 1 \\M &= 1 \quad S = -1 \text{ (phase inverted)}\end{aligned}$$

2. Boost/Cut Side channel level by applying gain $G \geq 0$ to the side channel signal

$$\begin{aligned}M &= 1 \\S &= G (2x - 1)\end{aligned}$$

For gain $0 \leq G < 1$, the side channel level is attenuated. For unity gain $G = 1$, the side channel is unaltered. And for gain $G > 1$, the side channel level is amplified.

3. Transform back to Left (L) and Right (R) channels:

$$\begin{aligned}L &= \frac{1}{2} (M + S) = \frac{1}{2} (1 + G (2x - 1)) = \frac{1}{2} + G [x - \frac{1}{2}] \\R &= \frac{1}{2} (M - S) = \frac{1}{2} (1 - G (2x - 1)) = \frac{1}{2} - G [x - \frac{1}{2}]\end{aligned}$$

Examples:

<u>Originally Center Pan:</u>	$x = \frac{1}{2}$
	$L = \frac{1}{2}$
	$R = \frac{1}{2}$

Therefore, Mid/Side processing with any value of gain G leaves center-panned signals unaffected.

<u>Originally Hard-Left Pan:</u>	$x = 1$
	$L = \frac{1}{2} + G/2$
	$R = \frac{1}{2} - G/2$

For $G = 0$ ($-\infty$ dB), the signal becomes center panned (mono).

For $G = 1$ (0 dB), the signal is unaffected and remains panned hard left.

For $G = 2$ (+6 dB), the signal is boosted in the left channel $L = 1 \frac{1}{2}$ and has a phase-inverted signal in the right channel $R = -\frac{1}{2}$. As discussed previously in the post on Sound Source Localization, this small, phase-inverted signal in the opposite speaker creates the apparent effect of “pushing” the sound further to the left than the originally hard-left-panned signal. This “over widening” of the sound comes at the price of potential phasing problems that are starting to arise between the stereo speakers.

For $G \gg 2$, $L \approx G/2$
 $R \approx -G/2$

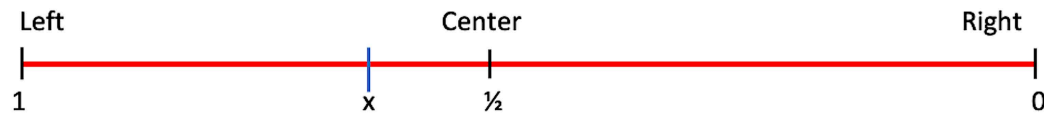
Here, the phase problem is fully evident – there will be “drop-outs” of this signal from the music !

<u>Originally Hard-Right Pan:</u>	$x = 0$
	$L = \frac{1}{2} - G/2$
	$R = \frac{1}{2} + G/2$

And the same comments above for the various values of gain G apply, with the words “left” (L) and “right” (R) interchanged.

Let's return to the issue of widening/narrowing the stereo image by using Mid/Side processing:

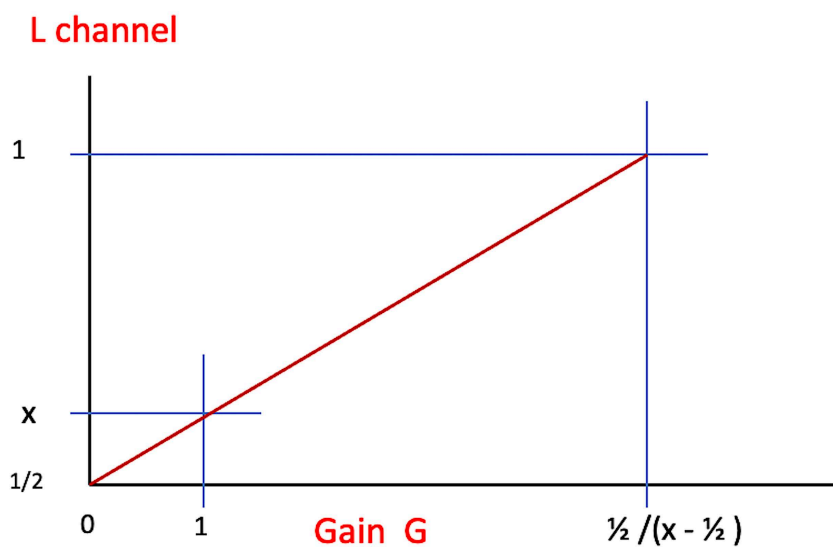
1. Left-Panned Signals



For signals panned to the left at a given value of x on the image ruler above, the left channel is given by

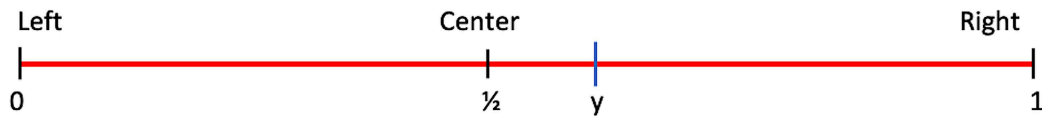
$$L = \frac{1}{2} + G [x - \frac{1}{2}] \quad \frac{1}{2} < x \leq 1$$

This left channel signal (L) is plotted versus gain (G) here:



If the gain G is less than 1 (attenuation), the signal is panned toward the center, thereby narrowing the stereo image. As the gain G is turned up above 1 (amplification), the signal is panned more towards the left speaker, thereby widening the stereo image. Note that as x approaches the value $\frac{1}{2}$ (center), no amount of practical gain can push the signal further to the left, *i.e.*, centered signals are unaffected by Mid/Side processing. Gain exceeding the value $\frac{1}{2} / (x - \frac{1}{2})$ should be avoided so that phasing problems between left and right channels don't occur.

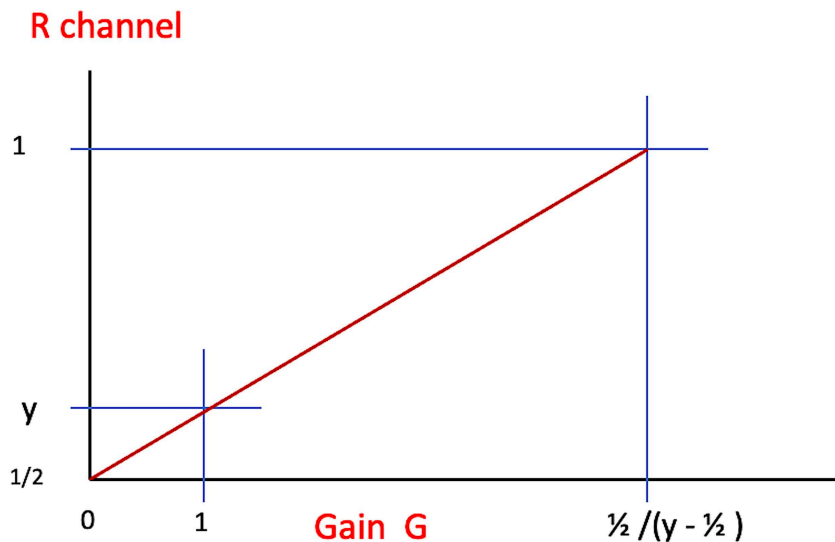
2. Right-Panned Signals



For signals panned to the right at a given value of $y = (1 - x)$ on the image ruler above, the right channel is given by

$$R = \frac{1}{2} + G [y - \frac{1}{2}] \quad \frac{1}{2} < y \leq 1$$

This right channel signal (R) is plotted versus gain (G) here:



Once again, if the gain G is less than 1 (attenuation), the signal is panned toward the center, thereby narrowing the stereo image. As the gain G is turned up above 1 (amplification), the signal is panned more towards the right speaker, thereby widening the stereo image. Note that as y approaches the value $\frac{1}{2}$ (center), no amount of practical gain can push the signal further to the right, *i.e.*, centered signals are unaffected by Mid/Side processing. Gain exceeding the value $\frac{1}{2} / (y - \frac{1}{2})$ should be avoided so that phasing problems between left and right channels don't occur.

The imaging plug-in on a stereo channel works to widen (and sometimes narrow) the stereo field in a number of different frequency bands by applying different amounts of gain to the side channel signal in each of these frequency bands. In a way, this is like having a broad graphic equalizer (EQ) in the side channel signal processing chain. Keeping the bass fairly focused near the center may actually require applying negative gain (in dB) in the bass frequency band to narrow its stereo image. Applying positive gain (in dB) in the higher frequency bands will widen the top-end and give the illusion of a wide mix.

C7. Stereo-ization



NaturalSoundsYEAH! , CC BY-SA 4.0, via Wikimedia Commons

Creating a powerful sense of space and presence in your recorded music is a primary function of the mixing process. The spatial balance in the mix is typically discussed in terms of “width” of the stereo image and “depth” of the acoustic listening environment. The former is primarily set by the Left (L) – Center (C) –Right (R) panning of the individual tracks to the Main Output stereo bus. The latter is primarily set by time-based effects, such as Reverb and Delay, that are added to the mix.

I’ve discussed previously many topics associated with creating and manipulating the stereo field image in chapters on **sound source localization**, **stereo image in mastering**, **mid/side processing**, and **stereo microphone techniques**. In this and the next chapter, I want to touch on the subject of the stereo imaging of individual audio tracks, versus the stereo imaging of the Left and Right channels of the main mix output bus. Here, I’ll look at “stereoizing” a mono audio track -- in other words, creating a wide, thick stereo sound from a thin, lifeless mono sound.

A commonly used approach for stereoizing a mono track involves utilizing the psychoacoustic phenomenon known as the **precedence effect**, discovered by Dr. Helmut Haas in 1949. Usually called the **Haas Effect**, it states that when one sound is followed by another with a delay time of approximately 40 milliseconds (ms) or less (below humans’ echo threshold), the two are perceived as a single sound. Neuroscientists believe that sound localization in the horizontal plane relies on two “cues”: the sound amplitude (loudness) difference between the two ears (inter-aural level difference ILD), and the time difference (delay) of sound reaching each ear (inter-aural time difference ITD). Two sounds with a short delay between them are perceived as one sound. The first arriving sound will localize the sound source, whereas the slightly delayed sound gives the human brain a sense of spaciousness, regardless of where this second sound comes from.

This **Haas Effect** can be used to create the perception of dimension to an otherwise “flat” mono audio track. This process of stereoizing a mono track goes as follows:

1. Duplicate a mono audio track,
2. Pan one track left and the other track right,
3. Add a short delay to one of the tracks.

Delays of 10 ms or less actually tend to *enhance* directionality, and yield “out-of-phase” type sounds that are hollow, thin, and sometimes dropping out completely. I’ve mentioned these phasing problems before when talking about **extending the stereo field beyond the speaker locations**. More examples of deleterious phasing effects are provided in the videos below. Delays between 10 ms and 40 ms, on the other hand, can produce the desired spaciousness. Choosing the right amount of delay is the trick. Ultimately, a stereoized track that sounds good in mono playback will ensure a solid translation regardless of the playback device and listening environment. A critical listening to the mix in mono serves as a quality assurance check for the sound. If phasing problems are present, this becomes clear when you hit the mono switch – the stereoized track shrinks and loses presence.

Joe Gilder is a songwriter, musician and sound engineer based in Nashville, TN. On YouTube, he offers a whole series of **expert tutorials on music production** using the **PreSonus Studio One** digital audio workstation. In the video below on stereoizing a mono track, he demonstrates what happens when you record a single input source (a guitar, for example) into a stereo track with the phase of one input inverted. In stereo playback of this track, there is a “push-pull” effect of the sound waves coming from the left and right speakers since one of the waveforms is an inverted version of the other. This is uncomfortable for the listener. And when the track is played back in mono, the sound disappears completely ! This happens because there is total cancellation of the two waveforms when summed. Now, mono playback is not as uncommon as you might think -- it effectively happens when listening to your phone speaker or in a large venue when you are at great distance from the left and right speakers. It can happen in your car. So, having this guitar part drop out completely is a big mistake.



OK, so what does Joe Gilder think about the stereoization technique of duplicating tracks with one channel slightly delayed from the other ? Take a look at this next video.



Well, the stereoized track sounds pretty good. But Joe is NOT happy with the mono playback -- there are phasing problems that lead to his hearing muddy bass tones and high-frequency 'metallic' sounds. Perhaps employing a greater delay in the 10 – 40 ms range might alleviate these problems ?

The goal of doubling a single source instrument part should be to obtain two separate, distinct waveforms, *i.e.*, decorrelated waveforms. This should minimize the unwanted waveform phasing effects and cancellations. In essence, doubling an instrument part should sound like making two different instrument parts. This can be done by inserting a splitter plug-in in the stereo track, separating the duplicated parts to left and right channels with different signal processing chains. Different waveforms can be achieved in each channel by inserting any or all of these signal processing plug-ins in each channel and using different parameter settings in the plug-ins:

1. Spectral processing – Equalization (EQ)
2. Dynamics processing – Compression
3. Gain
4. Time-based Effects (FX) – Delay
5. Modeling FX – amplifier

An example of this stereoization approach is shown in the following video. Joe Gilder records a single source guitar part to a stereo track, and inserts a channel splitter with two different amplifier plug-in sounds on the separated channels.



OK, Joe is now pleased with both the stereo and mono playbacks !

A Final Word

Since the goal of stereoization is to obtain two different waveforms of the same instrument part to form a stereo image, why not just do the following up front:

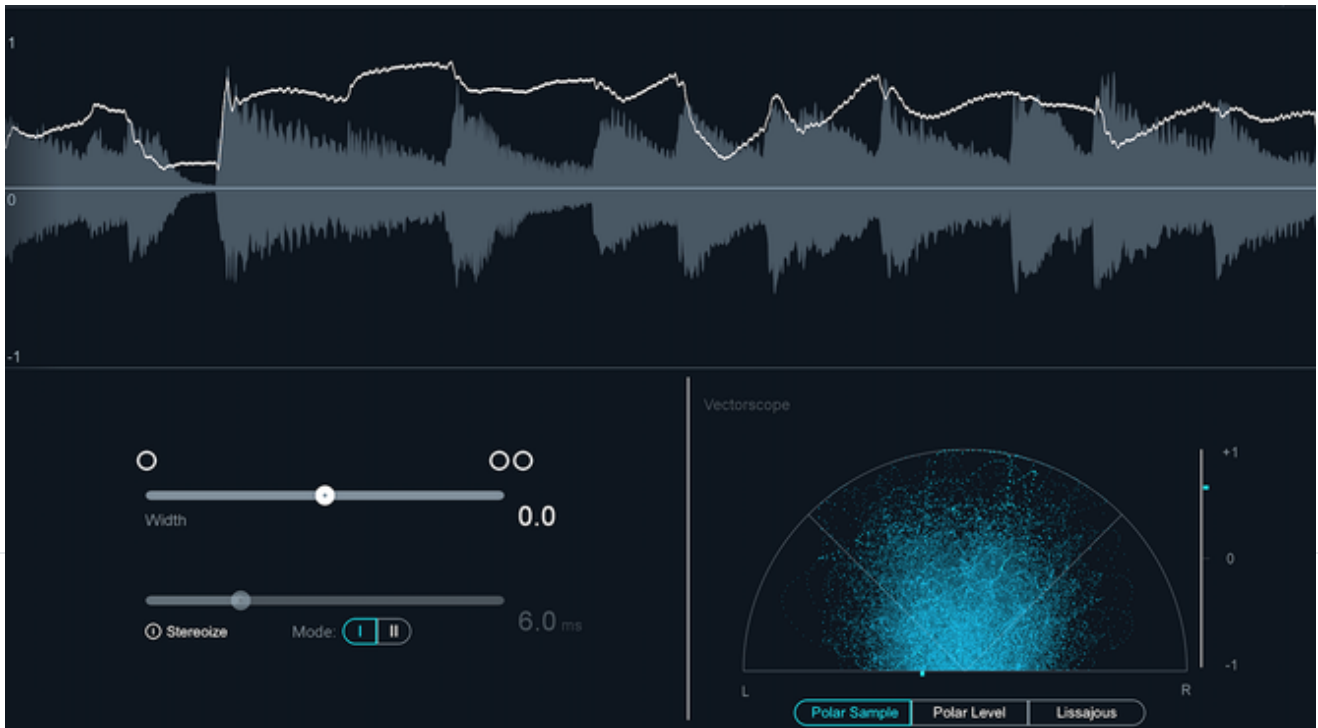
1. Use **stereo microphone techniques** to record to a stereo track in the first place, as shown in the lead photo at the top of this chapter,

OR

2. Actually record a second performance of that instrument part to a separate track -- this is truly doubling the part with two separate, distinct waveforms ! This approach is demonstrated very nicely in Joe's second video above (beginning at the 4:15 time stamp), and, in his expert opinion, is absolutely the best way to get a rich, spacious stereo sound in an audio track.

In the next chapter, I'll discuss several ways to manipulate the stereo image in an existing stereo audio track.

C8. Imaging a Stereo Track



Stereo Field Imaging - Phase Meter (click to enlarge)

In the previous chapter, I talked about **“stereo-izing”** a mono audio track. In this chapter, I’ll take a look at several ways to manipulate the stereo image in an existing stereo audio track. The imaging process is pretty much the same as what we do with a **stereo field imager** plug-in on the Left/Right channels of the Main Mix Output bus. This imager plug-in is shown in the photo above. But the goals of stereo field imaging on a single track are usually different from those of the Main Mix Output.

The first question is: what does the pan control knob in your mixer’s channel strip do on a stereo track ? The answer is not obvious. Once again, here’s **Joe Gilder of PreSonus** to walk us through the issue of panning on stereo tracks.



So, the answer to the question about the pan control knob is this: If you pan left, the left stereo channel output remains the same while ‘fading out’ the right stereo channel output. And vice-versa if you pan right.

Well, this certainly is not much ‘control’ over panning options. In fact, this does not seem particularly useful ! Fortunately, there are two very useful plug-ins that can give you full control over panning and stereo imaging. These two plug-ins are described in the second half of Joe’s video above. These plug-ins are:

1. Dual Pan

You get two separate pan control knobs – one for the left channel, one for the right channel. This gives you a lot of flexibility in changing the width and position of the stereo field on the track.

2. Binaural Pan

There are two major uses for this panning control :

- (a) Hit the ‘mono’ button, and use the pan control knob to move the combined (summed) L/R channels to the left or right, **OR**
- (b) Use the Width control knob to narrow (0 – 100%) or widen (100 – 200%) the stereo field. (Pan control knob remains at Center.) Below, a good, “one minute” YouTube video by Joe Gilder demonstrates this width control on a stereo track (two microphones) recording of a guitar part. Once again, there is the ever-present warning about widening the field too much and creating a “thin” sound from phasing problems.



Mid/Side processing is a very common technique for manipulating the stereo field image. Mid/Side (M/S) processing on a Left/Right (L/R) stereo track can be accomplished by placing the following Inserts in the mixer's channel strip :

- Mix Tool – MS Transform (L/R to M/S) (add +3 dB gain due to loss in the transform)
- Splitter (on Channel Split mode)
- Insert plug-ins (EQ, Comp, Gain) to each channel (Mid and Side)
- Mix Tool – MS Transform (M/S to L/R) (add +3 dB gain)

As we saw in the chapter on **Mid/Side processing**, increasing/decreasing the Side level gain with respect to Mid level will have the effect of widening/narrowing the stereo field. And using equalization (EQ), you can adjust the stereo width for individual frequency spectrum bands, e.g., narrowing the width of low-frequency sound while widening the width of high-frequency sound.

To wrap up this chapter, here is Joe Gilder to walk us through using M/S processing on a stereo track.



D. Mastering

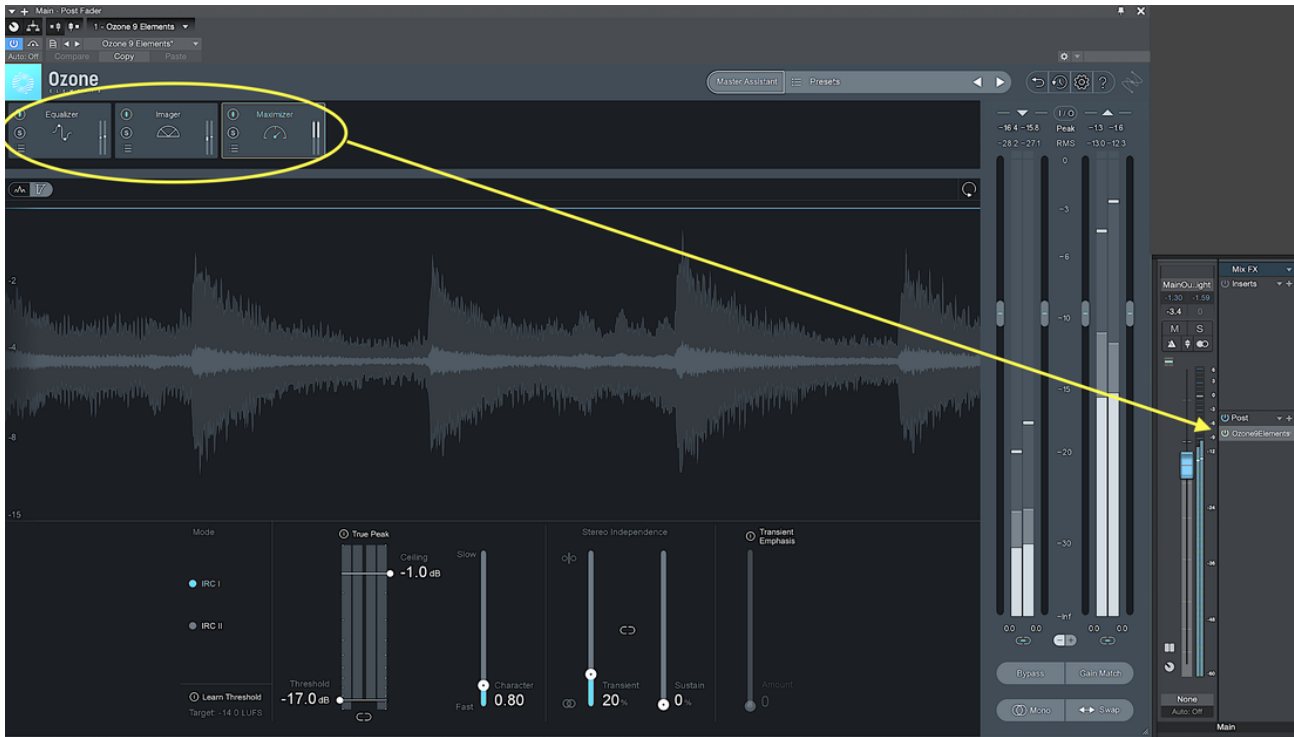
D1. Mastering - Overview



Mastering is the art of taking one or more mixed recordings (more precisely, the mixed stems from multi-track recordings) and shaping the overall musical qualities to achieve the desired sound to be put on the finished album product (CD, vinyl, mp3, m4a, WAV, etc.). Working with specialized audio gear and plug-in tools in digital audio workstations, a highly skilled mastering engineer with really good ears is usually tasked to do this. Typically, in home recording studios, both mixing and mastering is done by the same person – the artist herself/himself. Of course, this is not the ideal situation, as the artist is really “too close” to the music and lacks the objectivity (and frankly, the skill) to do the mastering. A second set of ears, with critical listening experience, would be preferable.

The goal of mastering is to achieve the desired overall balance in spectral tone, dynamics, and stereo imaging in each song and between all the songs in the album. For an individual song, mastering offers a chance to fix up or tweak the balances that were set during the mixing process, and to bring up the volume to the desired sound level. For a set of songs, mastering provides consistent sound balancing across the whole set, so that the album has a unified sonic character.

In my home studio, I rely on my **PreSonus Studio One** digital audio workstation (DAW) and the **iZotope Ozone 9 Mastering** software plug-ins to do the mastering. The Ozone 9 plug-ins are inserted post fader in the Main Output stereo channels of the DAW, as shown here.



[click on image to enlarge](#)

The EQ, Imager, and Maximizer plug-ins are used for tonal, stereo, and volume balancing, respectively. In the next three chapters, I'll review the function of each of these in the mastering process.

D2. Mastering - EQ



Mastering on Main Output (Mix) Bus (click image to enlarge)

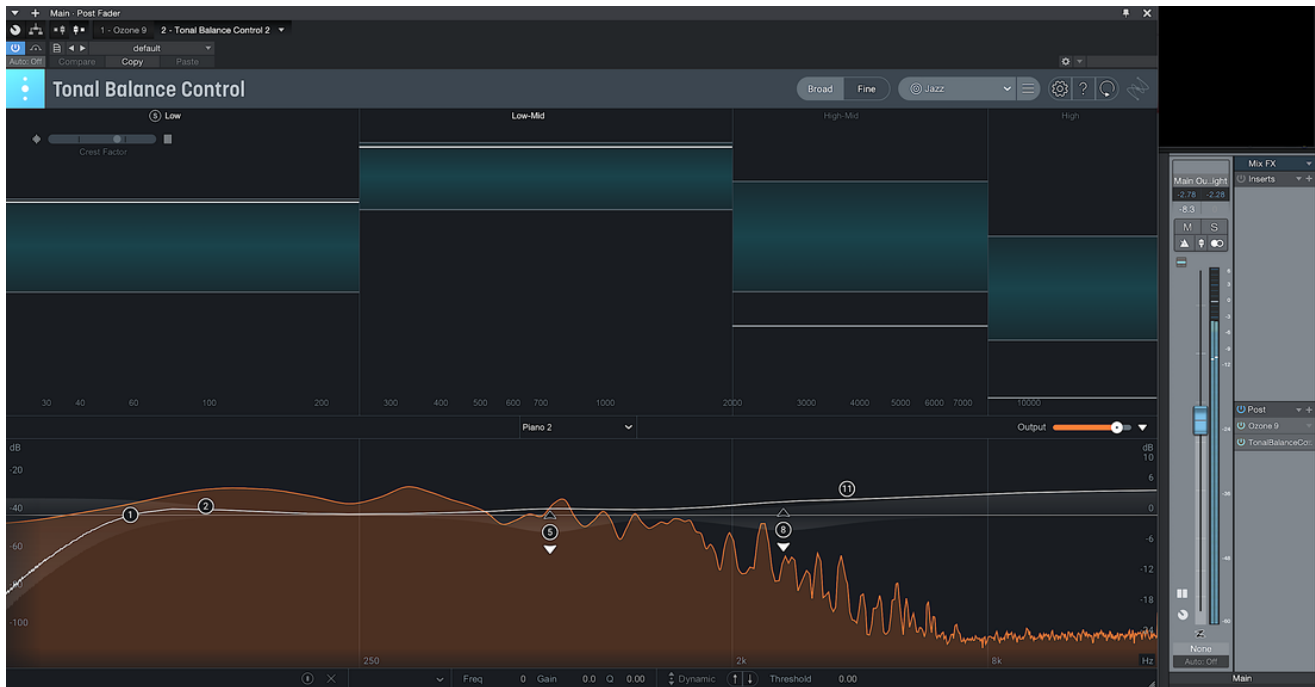
As discussed in the previous chapter Mastering Overview, the goal of mastering is to achieve the desired overall balance in spectral tone, dynamics, and stereo imaging in each song and between all the songs in the album. For an individual song, mastering offers a chance to fix up or tweak the balances that were set during the mixing process, and to bring up the volume to the desired sound level. For a set of songs, mastering provides consistent sound balancing across the whole set, so that the album has a unified sonic character.

In my home studio, I rely on my **PreSonus Studio One** digital audio workstation (DAW) and the **iZotope Ozone 9 Mastering software** plug-ins to do the mastering. The Ozone 9 plug-ins are inserted post fader in the Main Output stereo channels of the DAW. In this chapter, I'll take a look at the Master EQ, shown in the figure above as the first plug-in (furthest on the left) in the output chain.

Equalization (EQ) is applied to the whole stereo mix to give it the best tonal balance and to sculpt the overall sound. This is done to ensure that your song sounds great (and professional) to listeners. This is where the skilled ears of a mastering engineer come into play. Comparing your mix to a reference track that has the kind of sound that you are shooting for, the mastering engineer will make small broadband cuts/boosts in the frequency spectrum. The key thing to note here is that these EQ adjustments are “broad brush” and subtle – you don’t want to drastically change the sonic foundation of your mix.

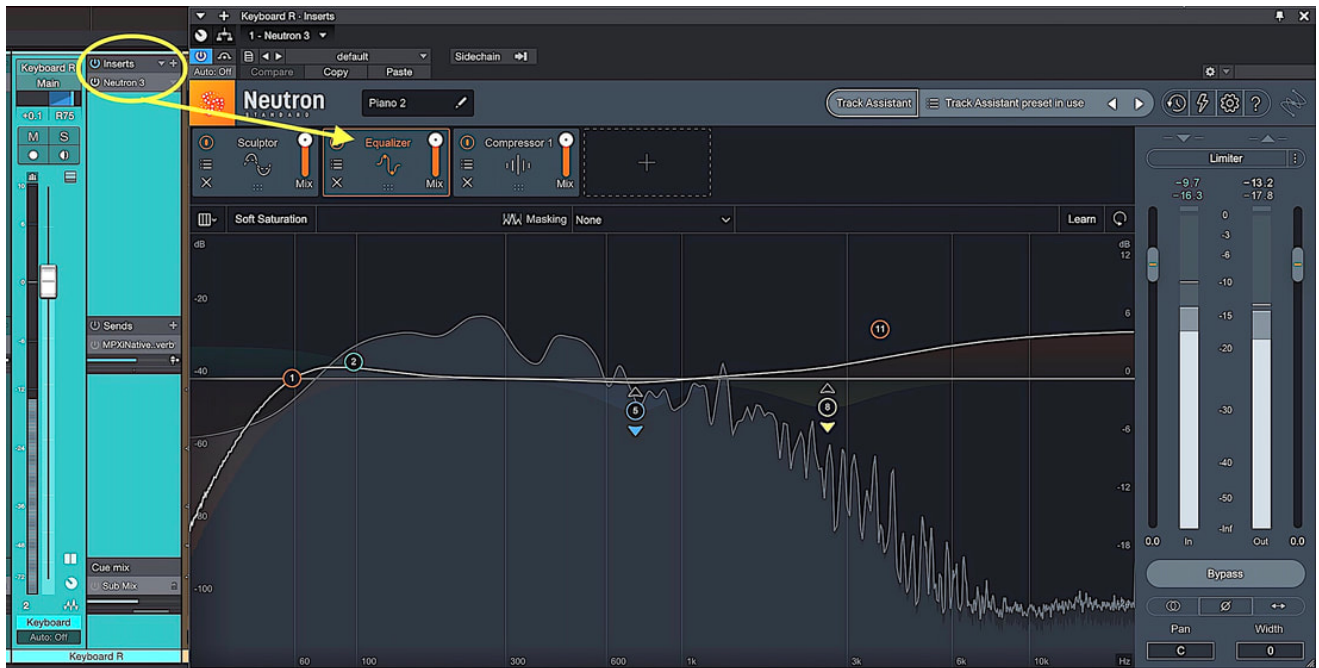
Typically, in home recording studios, both mixing and mastering is done by the same person – the artist herself/himself. Of course, this is not the ideal situation, as the artist is really “too close” to the music and lacks the objectivity (and frankly, the skill) to do the mastering. In the absence of a second set of ears to do the mastering, the iZotope Ozone 9 software offers two kinds of help – the Master Assistant and the Tonal Balance Control plug-in.

1. The Master Assistant is an AI (artificial intelligence)-powered tool that will make EQ and level adjustments to your tracks based on a preset reference music type (like “Classical” in the figure above) or on a custom reference track loaded into the software. In the example above, there are slight boosts in the low and high frequency bands, giving the sound an expansive and airy quality.
2. The **Tonal Balance Control** plug-in is the final one in the output chain. You can see it in the figure at the top as the last post-fader insert in the Main Output channels. Essentially, it is a frequency spectrum analyzer that compares the final tonal balance in your song to a preset reference genre or to a custom reference track loaded into the software. This capability is really nice, since this analysis is independent of your playback sound system and listening environment which are ‘coloring’ what you’re hearing. Using Tonal Balance Control can help assure proper tonal balances no matter what playback system the listener is using.



Tonal Balance Control (click on image to enlarge)

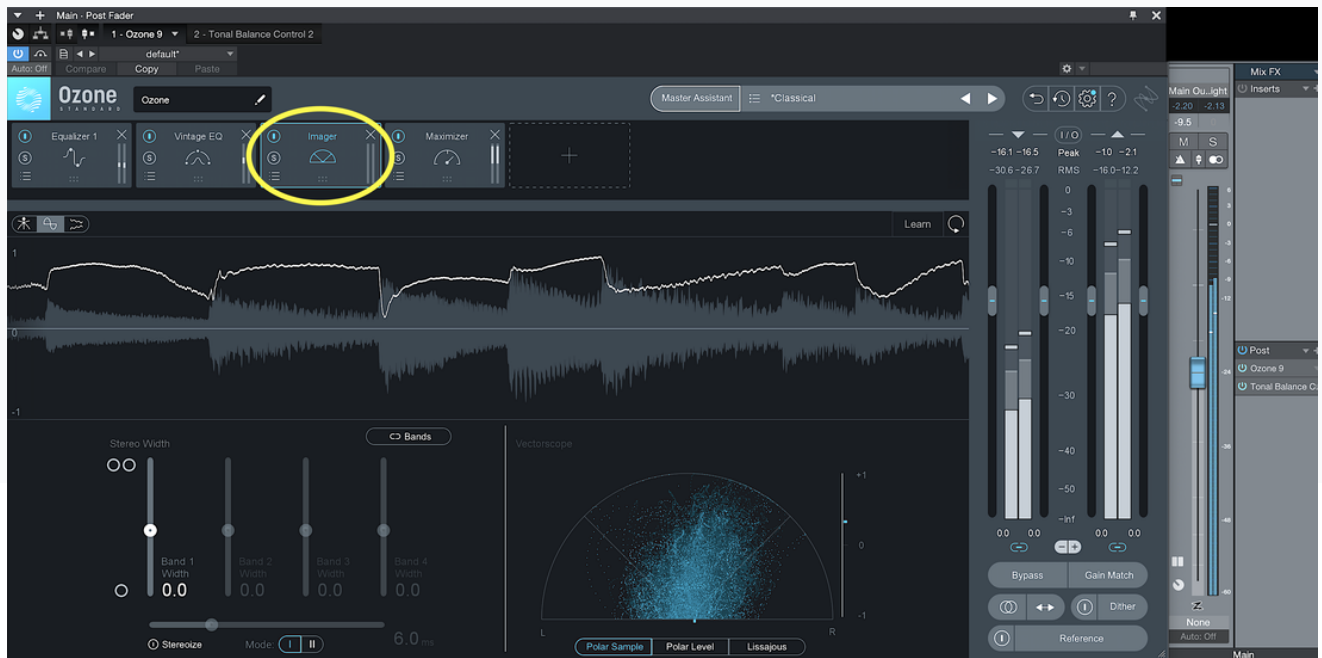
In this figure, low, low-mid, high-mid, and high frequency spectral levels are broadly viewed against targets based on a “Jazz” genre to identify potential trouble areas. By activating the “solo” function in a specific frequency band, you can hear what instruments are contributing the most to that part of the spectrum. A really cool feature is that you can adjust the **iZotope Ozone EQ** on the stereo mix bus AND the **iZotope Neutron EQ** on all of the individual channels in the mix directly from the Tonal Balance Control window ! So it's possible to make tweaks “on the fly” as you view the changes in real time against your reference targets. This communication between all the iZotope plug-ins is amazing, but you do need to have the ‘standard’ or ‘advanced’ versions of iZotope Ozone and Neutron to make this communication happen. (It doesn't work with the less expensive ‘elements’ versions.) As an example, the EQ ‘piano 2’ in the Tonal Balance Control window above communicates back to the Neutron EQ inserted in the Keyboard Right channel shown below. So, EQ changes made in the Tonal Balance Control window are automatically made back in the EQ channel insert. This really is a helpful feature for applying EQ in the mastering process.



Neutron EQ insert on channel strip (click on image to enlarge)

In the next chapter, we'll take a look at the next mastering plug-in in the output chain – the Imager.

D3. Mastering - Stereo Image



Imaging in the Mastering Process (click on image to enlarge)

The next mastering process in the chain of post-fader inserts in the Main Output stereo bus is "imaging". The **iZotope Ozone Imager** plug-in follows the EQ plug-in, as shown in the figure above.

The spatial balance in the mixed recording is typically discussed in terms of “width” of the stereo image and “depth” of the acoustic listening environment. The former is primarily set by the Left (L) – Center (C) – Right (R) panning of the individual tracks to the Main Output stereo bus. The latter is primarily set by time-based effects, such as Reverb and Delay, that are added to the mix.

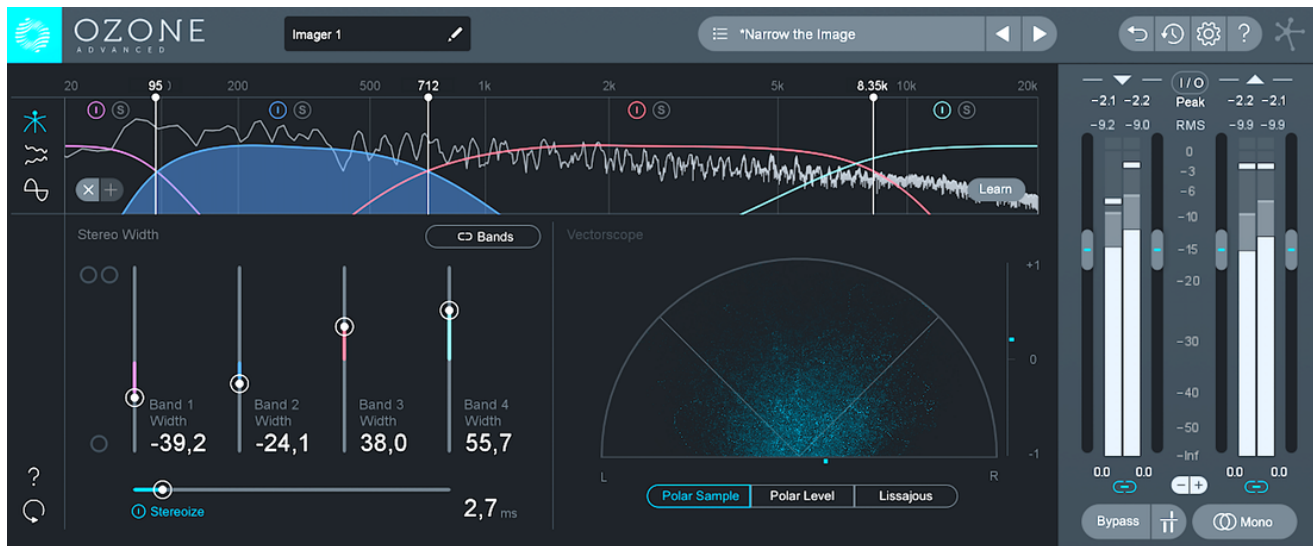
In recorded music, there has been an informal, generally held notion that the stereo image is a re-creation of a stage, an imaginary space stretching between the two speakers and reproducing the reality of live performances. In classical music recording, this notion has some validity, because the recording techniques used to create the recording usually include the use of a concert hall or stage, and the recording engineers make the evocation of that concert hall ambience a central part of their vision. However, such an approach is only one of many ways to use stereo to evoke a powerful sense of space and presence. For example, in popular music, the various recorded tracks can be panned according to a desired musical style:

Side		Mid		Side
Left		Center		Right
High Freq:		Lead Vocal		High Freq:
Guitar, KB		Solo		Doubled Guitar, KB
	←	Stereo Vocals, KBs	→	
	←	Stereo Drum Kit	→	
Harmony Vocals		Kick Drum	Harmony Vocals	
		Bass		

LCR Panning of Musical Instruments in the Stereo Image

In addition to panning to achieve a full spatial sound, there are techniques to “**stereo-ize**” mono tracks. Such techniques include doubling tracks (either double mic’ing or cleverly duplicating the track), applying EQ differently on left- and right-panned tracks, and using effects such as Reverb, Delay, Vocal Doubler, and Chorus on stereo buses.

The iZotope Ozone Imager on the stereo main output bus works to widen (and sometimes narrow) the stereo field in four different frequency bands by using “Mid/Side” signal processing. This **Mid/Side signal processing** was discussed previously for those interested in this more advanced technique in the recording industry.



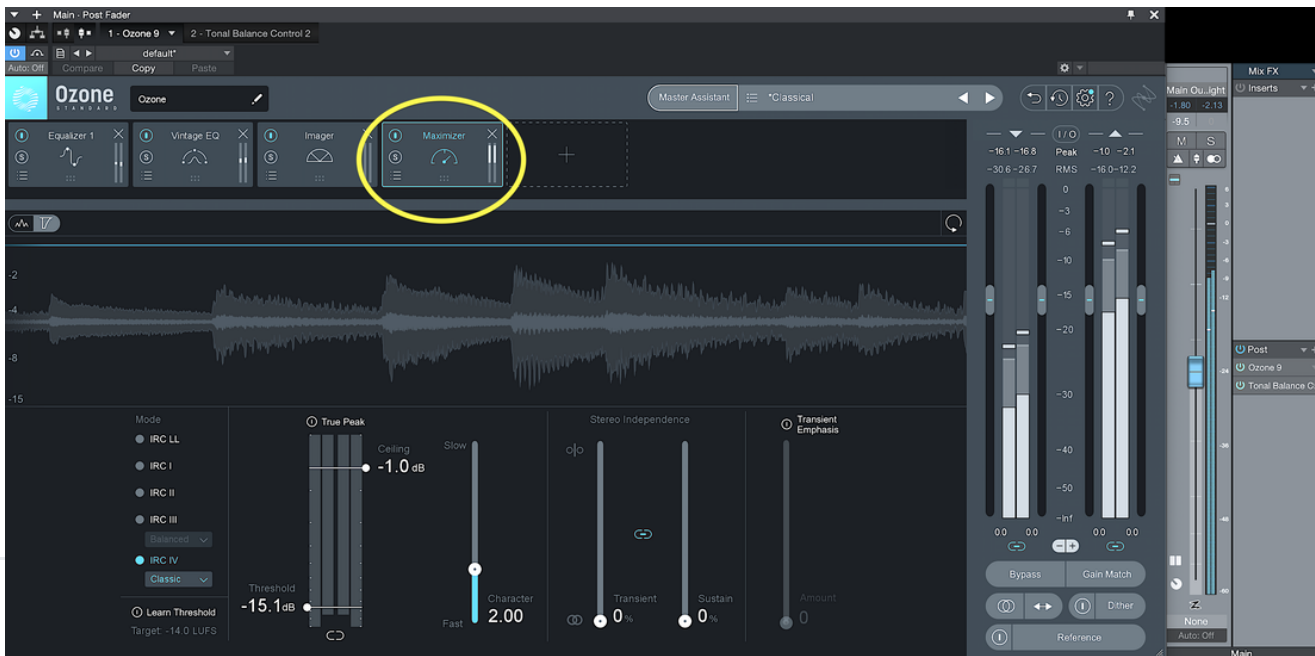
iZotope Ozone Stereo Imager Plug-in

The four adjustable frequency bands of Ozone's Imager offer flexibility when widening frequency-rich mixes. Keeping the bass fairly focused near the center may actually require applying negative gain (in dB) to the lower-frequency bands to narrow its stereo image. Applying positive gain (in dB) to the higher-frequency bands will widen the top-end and give the illusion of a wide mix. But like all things in music production, stereo image widening can easily be overdone, leading to disastrous results.

Just as you need a mix of dry and wet sounds to achieve mix depth and a combination of loud and soft elements for broad dynamic range, you need a spatial balance of narrow and wide signals for a mix to appear wide. Without any narrow signals, the listener has no point of reference for width. In fact, if every track is stretched to its limits, your mix will sound hollow and will suffer from phasing problems that will create drop-outs in the sound. Listener 'fatigue' will result from a loss in focus and clarity.

Ultimately, a mix that sounds good in mono will ensure a solid translation regardless of the playback device and listening environment. A critical listening to the mix in mono serves as a quality assurance check for the sound. If I've relied too much on panning and other widening techniques to expand the stereo image of my mix, this becomes clear when I hit the mono switch – phasing problems cause the whole song to shrink and lose presence. So, a very important caveat here: image processing must be applied sparingly and subtly ! But when used in just the right way, the imager can help give you an amazingly vibrant, spacious sound.

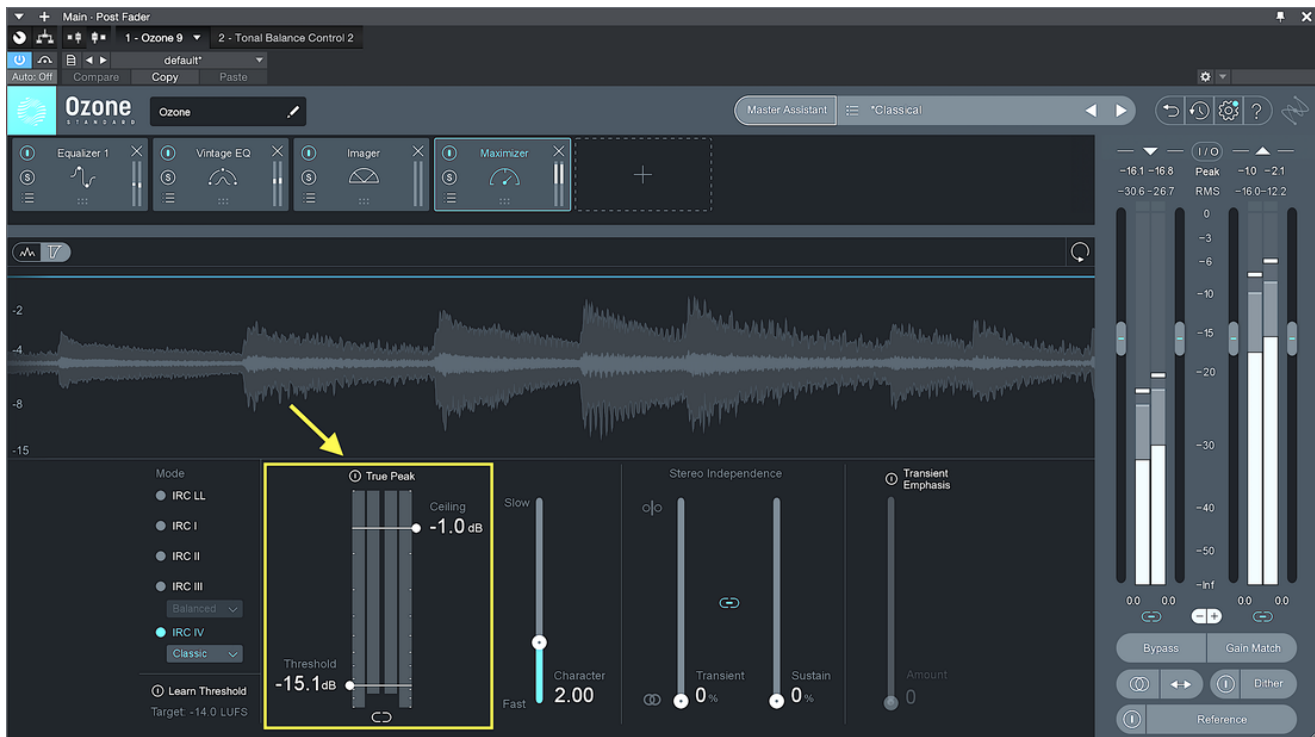
D4. Mastering - Limiting



Limiting in the Mastering Process (click on image to enlarge)

The final mastering process in the chain of post-fader inserts in the Main Output stereo bus is **loudness normalization** and **brickwall limiting**. This signal processing is done in the **iZotope Ozone Maximizer** plug-in, shown in the figure above.

The Maximizer is used to bring the average volume level of the mix up to a desired level while maintaining peak levels below the digital clipping level (0 dBFS). This process is known as “brickwall limiting” since it is like using a compressor with a very high compression ratio ($n \gg 10$) as a limiter. But the Maximizer is a special kind of limiter. It prevents the peak signal from going above a chosen ceiling level, typically in the range -1.0 to -0.3 dBFS. Therefore, the limiter protects against the terrible distortion caused by clipping of the digital signal. The plug-in uses “look ahead” samples to see peaks coming down the pike. And it can use inter-sample peak (ISP) metering to see “true” peaks.



True Peak Metering (click on image to enlarge)

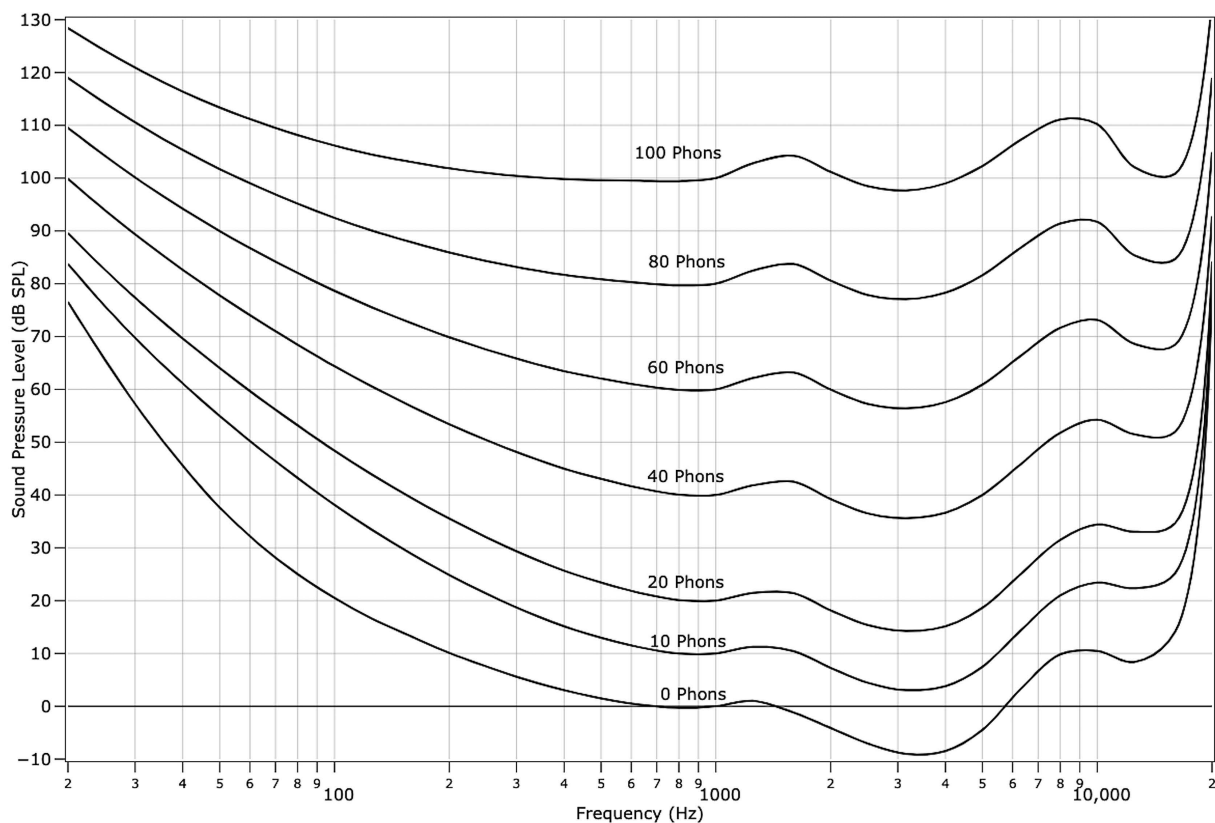
In the figure above, you can see that the ceiling level is set at -1.0 dBFS. The threshold level at which the limiting process engages is set at -15.1 dBFS. Since the threshold level also determines the amount of gain (dB) added to input level by the Maximizer plug-in, it has a significant effect on the overall average loudness and dynamic range of the song. In this example, the gain applied to the input signal is given by,

$$\begin{aligned} \text{Gain (dB)} &= \text{Ceiling (dBFS)} - \text{Threshold (dBFS)} = -1.0 \text{ dBFS} - (-15.1 \text{ dBFS}) \\ &= 14.1 \text{ dB} \end{aligned}$$

At this point, we need to say a few words about loudness “normalization” – adjusting your song loudness level to come close to reference (or target) levels chosen by the various music streaming services. These target levels are simply ‘goals’ for you to shoot for. A given music streaming service will “measure” the loudness of your song, and adjust it to its own common reference level. How a streaming service measures and adjusts your levels (the normalization process) is not standardized across the industry. As of mid 2021, Spotify uses International Telecommunication Union (ITU) Standard BS1770 to measure audio program loudness and true-peak audio level. Spotify’s target marks are time-integrated loudness level of -14 dB LUFS (see below) and true-peak (TP) level of -1 dBFS (often written as -1 dB TP).

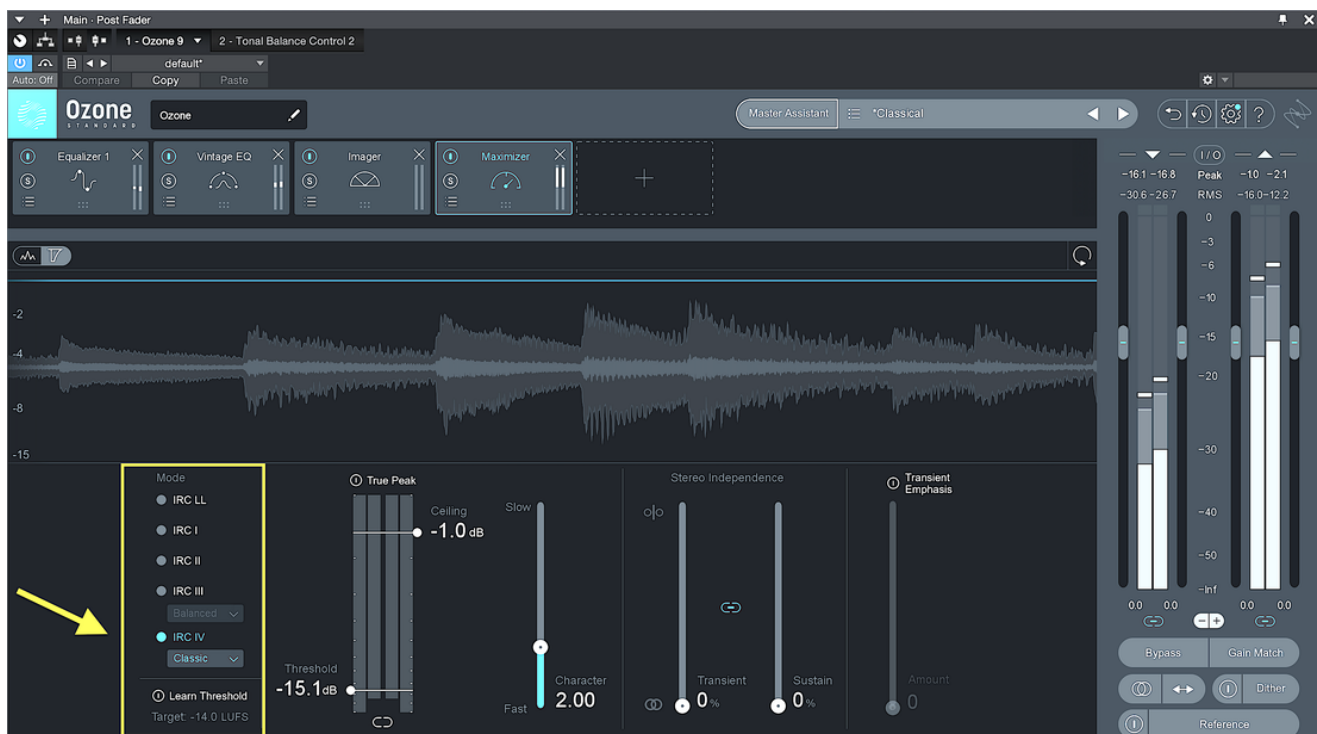
The “loudness” of a song can be measured by metering the amplitude of the signal waveforms. Average (RMS) and peak amplitude meters are most commonly used. This kind of loudness is a property of the signals only. But how about a measurement of the perceived loudness – what the human ear actually hears ? This requires quite a bit of knowledge of the perception part of human psychoacoustics, such as the equal loudness contours of the **Fletcher-Munson curves**. For each contour of constant loudness (units of Phons), you can see below that there is a wide variation in sound wave pressure level (dB SPL) across the frequency spectrum. For example, it takes a much higher sound wave pressure level (roughly +20 dB higher) to hear low-frequency sound at the same loudness as mid-frequency sound. The 0 Phons curve corresponds to the threshold of human hearing. The threshold of pain is approximately at 120 Phons.

Equal Loudness Contours (ISO 226-2003)



This is where the modern loudness measure “LUFS” comes in. LUFS stands for Loudness Units relative to Full Scale, and it’s based on the way our ears (and brains) react to the intensity of sound at different frequencies. With LUFS, mastering engineers can make a loudness measurement that takes everything into account. It’s the perception-based, time-integrated average loudness relative to the digital signal full-scale level of 0 dBFS.

Another important setting on the iZotope Ozone Maximizer is the selection of an “**Intelligent Release Control (IRC)**” mode of operation. In a previous chapter on **Compression**, the Release Time of the compressor was described. It is the time delay (in milliseconds) from when the signal level falls below the threshold to the de-activation of the compressor. An appropriate release time helps avoid ‘**pumping**’, the audible unnatural level changes associated primarily with the release of the compressor. A faster release behavior can result in a less noticeable pumping effect.



Intelligent Release Control of Limiter (click on image to enlarge)

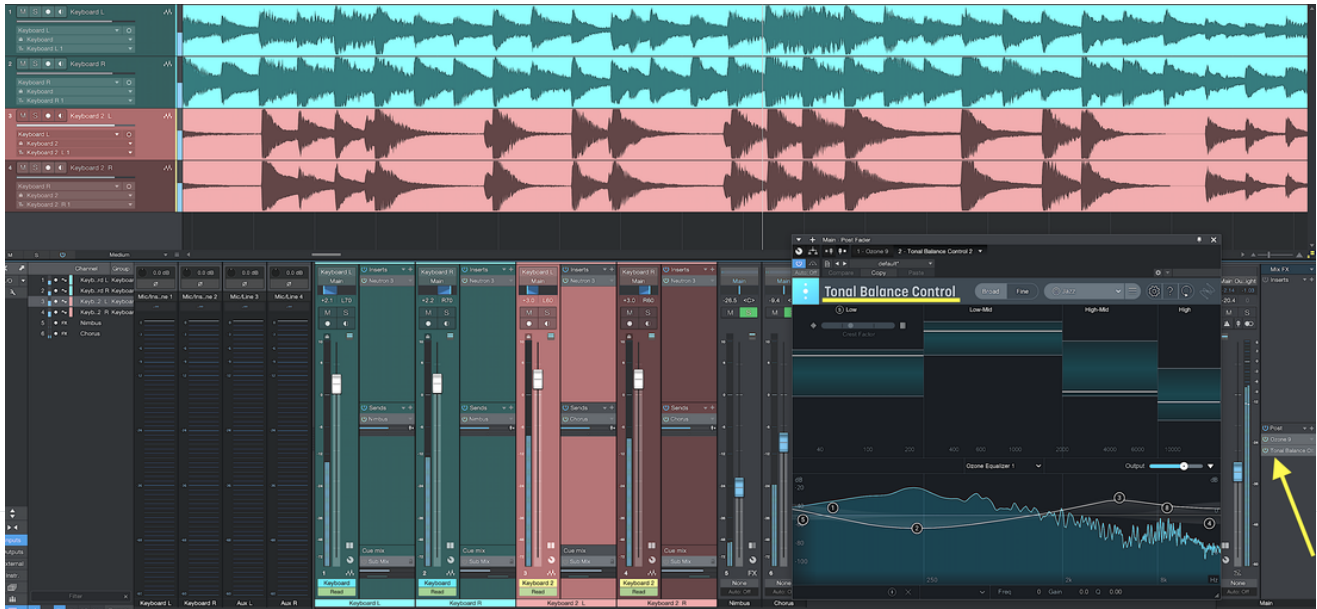
There are four IRC algorithms in the Maximizer plug-in that preserve the dynamics and overall clarity of your mix. **Mode IRC III** allows for aggressive limiting by using an advanced psychoacoustic model to intelligently determine the speed of limiting that can be applied

to the incoming signal, before producing distortion that is detectable to the human ear.

Mode IRC IV builds upon the IRC III technology by shaping the spectrum to further reduce pumping and distortion. As the signal level goes farther above the threshold level, the IRC IV algorithm limits frequency bands that contribute most to these peaks. Dozens of psychoacoustically-spaced frequency bands are utilized in order to achieve a more natural and transparent effect. The IRC III and IRC IV algorithms are very CPU-intensive, and produce high latency, especially at higher sampling rates.

Finally, the overall response time (attack and release times) of the limiter processing can be tweaked using the Character slider bar shown in the figure above. The actual attack and release times used are dependent on the selected IRC Mode. A continuous range from Fast (0.0) to Slow (10.0) is available in each mode.

D5. Mastering - Tonal Balance Control



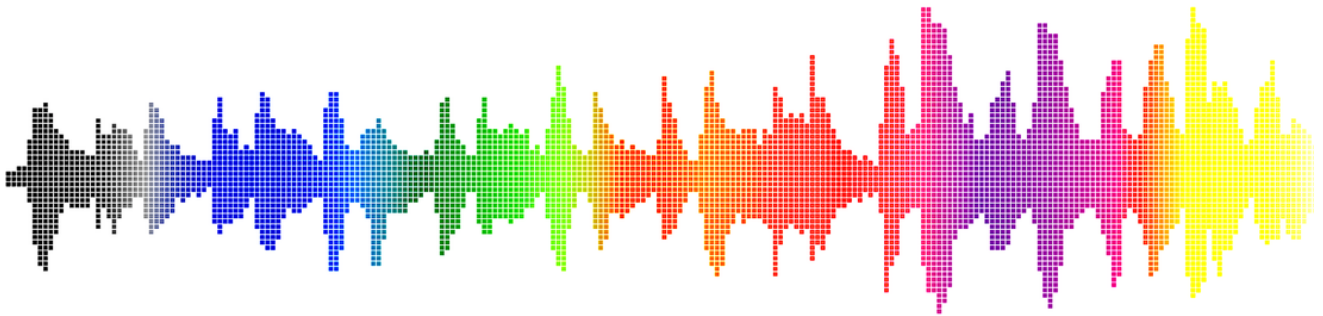
Tonal Balance Control in the Mastering Process (click on image to enlarge)

In the figure above, the **iZotope Tonal Balance Control** plug-in is the final one in the post-fader output chain of the Main Output stereo bus. It was introduced in a previous chapter as a visual aid that serves as a second pair of “ears” in the **Mastering EQ process**. Essentially, it is a frequency spectrum analyzer that compares the final tonal balance in your song to a preset reference genre or to a custom reference track loaded into the software. This capability is really nice, since this analysis is independent of your playback sound system and listening environment that may be ‘coloring’ what you’re hearing. Using Tonal Balance Control can help assure proper tonal balances no matter what playback system the listener will be using.

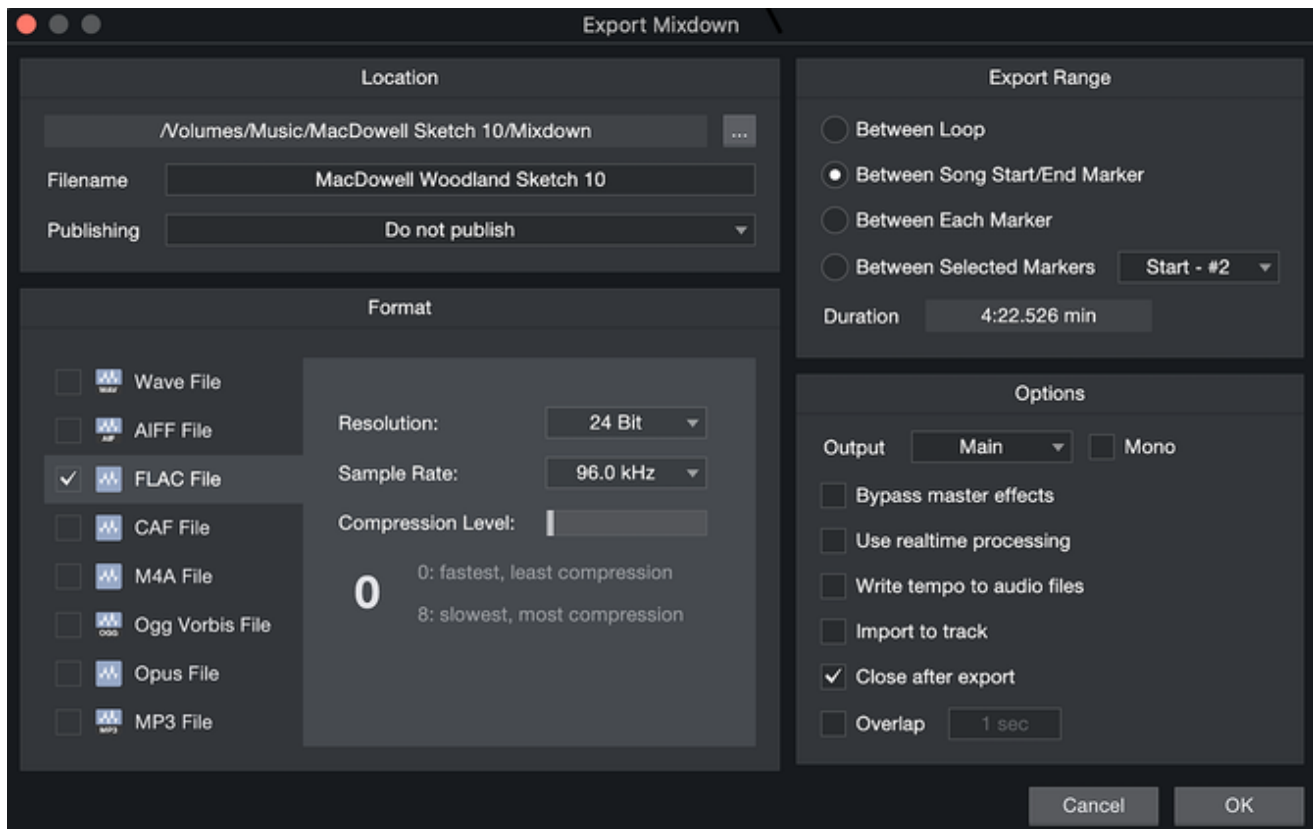
Broadband low, low-mid, high-mid, and high frequency spectral levels are viewed against targets based on a chosen genre (“Jazz” in the example above) to identify potential trouble areas. By activating the “solo” function in a specific frequency band, you can hear what instruments are contributing the most to that part of the spectrum. A really cool feature is that you can adjust the **iZotope Ozone EQ** on the stereo mix bus AND the **iZotope Neutron EQ and Relay gain** on all of the individual tracks in the mix directly from the Tonal Balance Control window ! So it's possible to make tweaks “on the fly” as you view the changes in real time against your reference targets. This communication between all the iZotope plug-ins is amazing, but you do need to have the ‘standard’ or ‘advanced’ versions of iZotope Ozone and Neutron to make this communication happen (it doesn't work with the less expensive ‘elements’ versions.) In the example above, the Ozone Equalizer 1 shown in the Tonal Balance Control window communicates back to the Ozone Mastering EQ in the Main Output stereo bus. So, EQ changes made in the Tonal Balance Control window are automatically made back in the Ozone Mastering EQ. The other EQs in the track channels are accessible via the drop-down menu in the EQ selection box in the Tonal Balance Control window. This feature really is helpful for applying EQ in the channels and buses during the mastering process.

In the next chapter, we'll take a look at what comes after the Mastering process – the exporting of the mixdown to final audio files.

D6. Export Mixdown



When your mixing and mastering work in the digital audio workstation (DAW) is completed, it's time to export (save) your stereo mix to a digital audio file for distribution online or on a CD. To create an audio file in the **PreSonus Studio One** DAW, the "export mixdown" menu is opened:



For your audio file, you can choose from a number of different digital formats. These formats can be categorized into three distinct groups: Uncompressed, Lossless, and Lossy. These three groups are summarized below.

Uncompressed

Uncompressed audio files are associated with professional quality sound. They have a bit depth and sampling rate of at least 16 bits and 44.1 kHz, respectively. (16 bits/44.1 kHz are CD quality settings.) Therefore, you can expect at least a dynamic range of 96 dB and a frequency bandwidth of 22 kHz. Because these files are not compressed in any way, *i.e.*, all the musical information data is kept intact, they accurately represent the master that was created. Of course, this also means song file sizes are very large. The two best formats in this category are:

1. WAV

The Wave (WAV) uncompressed file format is the “go-to” file type for professional recording, mixing, and mastering. Commonly used bit depths and sampling rates are 16, 24, or 32 bits and 48, 96, or 192 kHz, in various combinations. Apple Music, Spotify, Jamendo, and many other streaming services support and sometimes even demand WAV file uploads.

2. AIFF

The Audio Interchange File Format (AIFF) uncompressed file format is used on Apple's software. Like WAV formatted files, AIFF formatted files give the highest quality audio. The AIFF format was once exclusive to Apple operating systems, but now it will play on PC operating systems too.

Lossless

Lossless formatted files take up significantly less storage space than uncompressed files, and ideally retain full audio quality. Lossless formatting works by compartmentalizing redundant or repeated data, while providing a "set of instructions" for these parts to be recreated during playback. As a result, lossless files are about 50% smaller in size than uncompressed files, but still offer comparable sound quality during playback. The two best formats in this category are:

1. FLAC

The Free Lossless Audio Codec (FLAC) lossless format offers the highest quality to file size ratio of just about any other file format. It can recreate musical data as large as a 32 bit depth/192kHz sampling rate while being able to reduce file size by up to 70% in some cases ! Although FLAC is not supported by iTunes, it is supported by both Apple and PC operating systems.

2. ALAC

The Apple Lossless Audio Codec (ALAC) is Apple's implementation of lossless data compression of digital music. ALAC encoded files have M4A filename extensions. ALAC supports audio at 16, 24, and 32 bit depths with a maximum sample rate of 384 kHz. Audio files compressed with ALAC are reduced in size by 40-60% . ALAC is supported by Apple Music.

Lossy

Lossy formatted files are highly compressed files that produce small file sizes. They are created by identifying and then deleting data information that is mostly indiscernible by the average listener. This is a great option for streaming, or any online service in which speed of download is more important than the quality of the audio. The sound quality of lossy files can vary greatly – from high sound quality approaching lossless files, to low sound quality with noticeable aliasing artifacts, quantization distortion, and attenuated high-frequency energy. The two most popular formats in this category are:

1. MP3

The most popular audio file format in use today is the ISO-MPEG Audio Level-2 Layer-3, commonly referred to as MP3. The MP3 codec compresses audio by a substantial factor, leading to very small file sizes. A wide range of compression rates can be chosen to encode/decode an MP3 file – 128, 160, 192, and 320 kbps. The 320 kbps rate offers “near CD” sound quality, with the obvious trade-off being a larger file size.

Regardless of platform, operating system, or software, an MP3 will most likely play – making it a great choice for those looking to have their music instantly playable. If you’re willing to sacrifice quality for small file size, and want ease of use and streaming capability, then MP3 is a good file format for you.

2. AAC

Advanced Audio Coding (AAC) is the lossy compression codec intended to replace MP3. AAC encoded audio files have AAC and M4A filename extensions. AAC employs a much more complex compression algorithm than MP3 for removing less important data information. This means that AAC offers a higher sound quality than MP3 at very similar file sizes. Since AAC can be played on almost as many platforms as MP3, it is considered the best lossy file format currently available. AAC is supported by Apple Music.

When I export a mixdown of a song, I do so creating several audio files with different formats. These audio files are destined for uploads to different hosting sites. Here's a table listing the audio files of my piano recording of Edward MacDowell's Woodland Sketch Op. 51, No. 10. The duration of the song is 4 minutes, 22 seconds.

File	Format	Resolution	File Size	Host Site
Song.WAV	Uncompressed WAVE	24 bit 96 KHz	151.2 MB	Internet Archive archive.org/details/@pedalpointsound
Song.WAV	Uncompressed WAVE	24 bit 48 KHz	75.6 MB	Jamendo Music jamendo.com/artist/529848/gregory-tait
Song.FLAC	Lossless FLAC	24 bit 96 KHz Compression Level: 0 (least)	92.9 MB	MusOpen musopen.org/music/performer/greg-tait
Song.M4A	Lossless ALAC	24 bit 96 KHz	90.7 MB	Song download from Pedal Point Sound website pedalpointsound.com
Song.M4A	Lossless ALAC	24 bit 48 KHz	45.6 MB	
Song.MP3	Lossy MP3	48 KHz sample rate 320 kbps bit rate	10.5 MB	Song playback from Pedal Point Sound website pedalpointsound.com

Apple Music will play the WAV, M4A, and MP3 files. QuickTime will play the FLAC file.

E. Recording Project

E1. Session 1: Recording



In the next four chapters, I'll provide an overview of the four studio sessions for creating a musical recording that will be posted on a streaming platform. The four sessions are recording, mixing, mastering, and streaming.

In this first session, I'll cover the recording of Johann Sebastian Bach's Adagio in D minor for solo keyboard. This piece is very popular, and has been used in the soundtracks of several films, including *Coda* (2020) and *Fifty Shades of Grey* (2015). This Baroque-era work is emotional, meditative, even a bit dark, with a beautiful upper melodic line undergoing many stylistic ornamentations over a slowly plodding, steady progression of chords in the low-mid register. The Adagio is a solo keyboard arrangement of the second movement of Bach's Concerto in D minor, BWV 974, which in turn is a transcription of the Concerto for Oboe and Strings in D minor composed by Alessandro Marcello. Bach's autograph of this keyboard arrangement has been lost, but the arrangement was copied around 1715 by Bach's second cousin, Johann Bernhard. It is this copy that we use today to perform the Adagio.

Musescore.

BWV 974 - Adagio

Bach / Marcello

Please rate this score ?

☆☆☆☆☆

Difficulty Level Beta

I agree

I disagree

Intermediate

We rate difficulty level of this score as a Intermediate one.
Do you agree with that?

This score is based on

Edit

Bwv 974 - Adagio In D Minor

by Johann Sebastian Bach

Other versions of this composition

BWV 974 - Adagio in D minor – ★★★★★ 7 votes

Johann Sebastian Bach

Cello, Piano

bwv 974 g

Guitar (2)

Concerto in D Minor, BWV 974

Solo Harpsichord

[Show more](#)

Uploaded on Jun 18, 2012

Pages 3

Duration 03:06

Measures 41

Key F major, D minor

Genre Classical

Ensemble

Parts

Part 1

[Privacy](#)

License

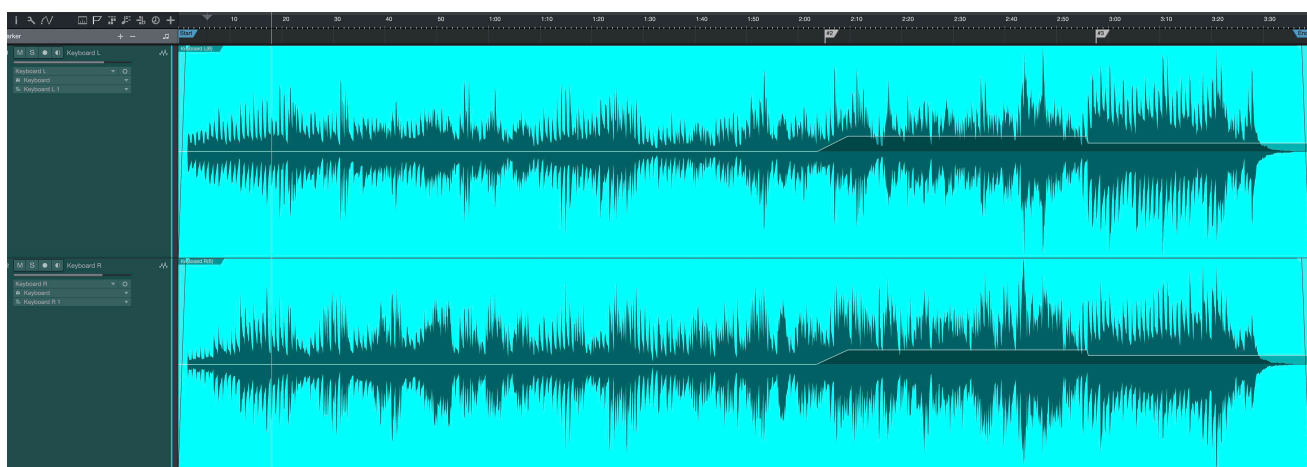
🌐 Everyone can see this score

 Creative Commons copyright waiver

Click on image to enlarge

Technically, this is not a difficult piece to play. But the artistry is in the articulation and expression of the ornamental melodic line, and in the nuanced tempo variations. This is why a 'live' performance of this work is so important. A MIDI-programmed virtual-instrument grand piano cannot fully capture the needed artistry, even with its articulation control, velocity profiles, and quantization and timing editing.

I recorded the piece in five ‘takes’ and used an editing process called “**comping**”, short for “compositing”, to splice together the four best sections from the takes. Two sections are joined together using a painstaking process of aligning sound waveforms from both sections and applying amplitude cross-fading. If done properly, there will be a smooth transition in timing and dynamics between the two sections, with no audible clicks ! This splicing has to be successful musically as well as acoustically. Since the left and right piano tracks are the only ones playing, an imperfect splice cannot be masked by the sounds of other instrument tracks playing. Lastly, the gain (loudness) profile of the recorded waveforms in both left and right channels is very slightly edited to achieve the desired dynamic character in the overall recording. I prefer to use nondestructive “**clip gain**” editing of the recorded waveforms, rather than volume automation of the channel faders during the mixing process. The final recorded sound waveforms are shown below in the **PreSonus Studio One Pro** digital audio workstation.



Recorded Sound Waveforms (click on image to enlarge)

In a previous chapter, I talked a bit about the philosophy of **editing recorded waveforms**. I want to reiterate some of those comments here, as I think it is worth being reminded of the ‘goal’ of editing.

When the recording “red light” comes on, it’s only natural to feel ‘nervous’ , just like you would feel performing the music in front of a live audience. You’re thinking: “This is it. What I record into the digital audio file is permanent. I’ve got to “get it” just the way I want, so that listeners will hear and enjoy a good musical performance.”

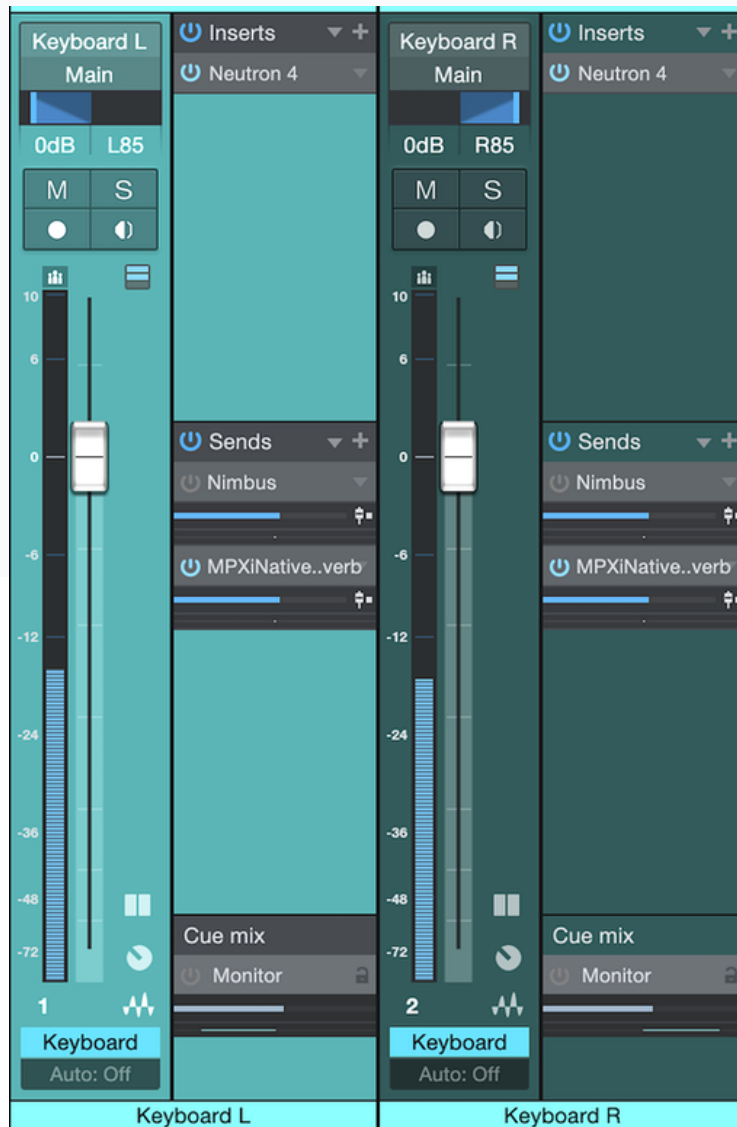
This is why it is necessary to have done all the hard work of practicing the piece to a sufficient level of technical mastery beforehand. And then, take a lot of time to allow your playing to develop and mature the musical ideas you wish to convey. Now, you can concentrate on the interpretation and musicality of the piece while recording the performance.

It is unreasonable, however, to expect to be able to “lay down” your very best performance in a single complete “take” . It is very comforting to know that you can, and will, use some editing of the recording, during the recording session as well as afterwards, to put together a performance you are happy with. Here are some excellent thoughts about the philosophy of editing expressed by the pianist **Paul Cantrell** on his website:

"An aside on the philosophy of splicing: I'm skeptical of classical recordings with hundreds of splices that are pieced together entirely in the studio — it works well when the studio is part of the compositional process, but when the purpose of such splicing is obsessively perfectionist correction of live playing, it is dangerous. It can lead to a clinically perfect but spiritless recording without the organic expressiveness that makes music magical. *However*, neither am I in the camp that decries splicing as some kind of hoodwink or moral failing; recordings are recordings, not live performances, and a musician's job is to work their medium's full potential to produce the best possible experience. So, when I splice, I try to strike a balance between correcting really conspicuous mistakes that disturb the flow of the music, and preserving that flow in its natural form. In short: splices are artistic decisions, and must be treated as an aspect of musical performance, not as cosmetic surgery. End of aside."

In the next chapter, I'll talk about the mixing session in the creation of the musical recording of Bach's Adagio in D minor for keyboard solo.

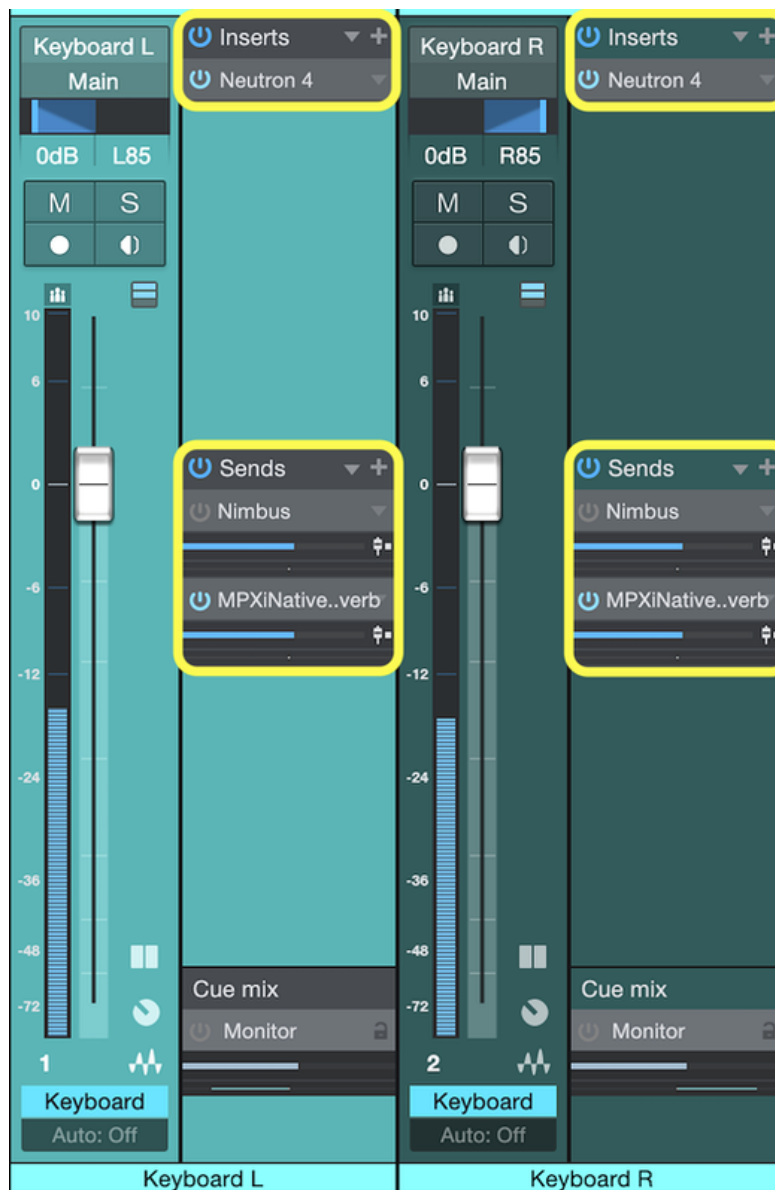
E2. Session 2: Mixing



In this session, the second of four, I'll talk about the mixing process in the creation of the musical recording of Bach's Adagio in D minor for keyboard solo. There are just two tracks in this recording – the left and right outputs of the digital piano. Rather than recording the piano to a single stereo track, I obtain more flexibility to process the left and right channel signals independently by recording to two mono tracks. The mixing console for these two channels is shown in the figure above.

The input gains of the audio interface amplifiers have been set and the input gains of the digital audio workstation (DAW) have been trimmed to achieve the proper volume balance between the left and right channels in the recorded waveforms. Therefore, there is no need to adjust the volume faders of both channel strips, and they are set at 0 dB. For imaging the signals to the stereo Main Output bus, I have set the left channel to 85% pan left and the right channel to 85% pan right. I don't pan the left and right channels to hard left and hard right, respectively. This is because I will be using some additional stereo field imaging in the mastering process – more about this in the next chapter.

The most artistic part of the mixing process comes from the signal processing that is applied to each channel. Equalization and dynamics processing plug-ins from the **iZotope Neutron 4** suite of professional mixing tools are included in the **Insert sections** of the channel strips. And the time-based effects (Reverb) plug-in from **Lexicon** is included in the **Sends sections** of the channel strips. These Insert and Sends sections can be seen below in the channel strips of the DAW mixer console.



Inserts

The **iZotope Neutron 4** mothership plug-in comprises a suite of signal processing plug-ins combined with artificial intelligence (AI) assistive audio technology for crafting a professional mix. iZotope is one of the, if not *the*, leading signal processing software developers in the music production industry today. It is the “go to” source for many professional music studios, and I have come to appreciate how amazing these plug-ins are in a home music studio too.

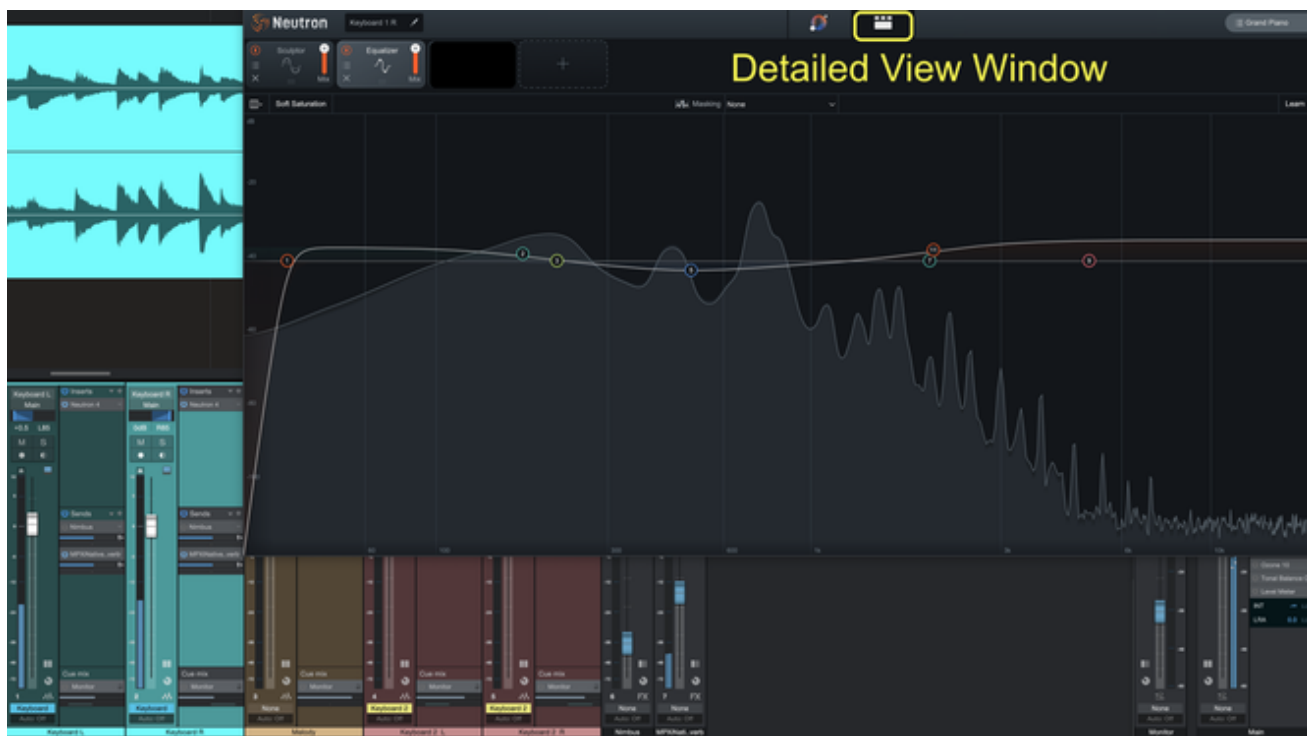
The process begins with invoking the AI-powered Assistant to make a “first cut” putting together a signal processing chain of plug-ins based on your desired sound intent for tonal and dynamic balance. This is shown below in the Assistant View Window of the Neutron 4 insert.



[Click on image to enlarge](#)

The user sets the intent controls of Tone Match, Punch, Distort, and Width, and the AI Assistant configures a chain of signal processing plug-ins that works toward achieving the desired tonal, dynamic, character, and width targets. The Punch, Distort, and Width intent controls are connected to Compressor (affecting signal transients), Exciter (creating harmonic distortion), and Imager (adjusting stereo field width) plug-in modules, respectively. I don't need these for my Bach Adagio piano tracks. But the **Tone Match** intent control *is* useful to me. The Tone Match slider controls two parameters simultaneously: (1) the Assistant Pre-EQ control in the Neutron mothership plug-in, and (2) the Amount control found in the Sculptor plug-in module. Both controls work together to shape the input signal towards the selected tonal balance target or custom reference curve in the Target Library. Note that the EQ and Sculptor plug-ins can shape the frequency spectrum content only where there is such content present in the signal energy of the piano tracks.

Once the Assistant has created a “starting point” in your mix, you can switch over to the Detailed View Window to see and make adjustments in the various plug-in modules.



[Click on image to enlarge](#)

Sculptor Plug-in Module

The first plug-in is the **iZotope Sculptor module**, which is a very powerful signal processing tool. It is described by iZotope as follows:

“Sculptor is a spectral-shaping tool that brings clarity and polish to your tracks. This means removing muddiness, reducing harshness, and helping shape your tracks into better versions of themselves. Spectral shaping is **multiband compression** taken to the extreme. Instead of compressing four frequency bands, spectral shaping compresses the signal in up to 32 frequency bands, allowing for a control that is more tailored for the signal. Compression thresholds can be set toward a desired spectral shape or remain adaptively adjustable to compress the signal “toward itself”, i.e. its own time-averaged spectral shape.

Sculptor will apply dynamics processing to areas within the frequency spectrum based on a threshold aimed at hitting the selected target curve. The target curves are the idealized spectral version of the selected instrument. Sculptor is designed specifically not to add any distortion, preserving the integrity of the original signal as transparently as possible.”

The topic of **Compression** – what it is and why it is used – has been covered in a previous chapter. The Sculptor multiband compression module that is used on the piano tracks in this recording is shown below.



[Click on image to enlarge](#)

EQ Plug-in Module

The second plug-in is a static Equalization (EQ) module. The topic of **Equalization** – what it is and why it is used – has also been covered in a previous chapter. The EQ module used on the piano tracks in this recording is shown below. Spectral refinements appropriate for a grand piano include a sub-bass roll-off, low-frequency boost, low-mid frequency cut to remove muddiness, and high-frequency boost to add clarity and airiness.



[Click on image to enlarge](#)

Sends

Reverb FX

Time-based effects (Reverb) is added to the mix using the classic **Lexicon MPX-i Native Reverb** plug-in. This Reverb FX is shown below.

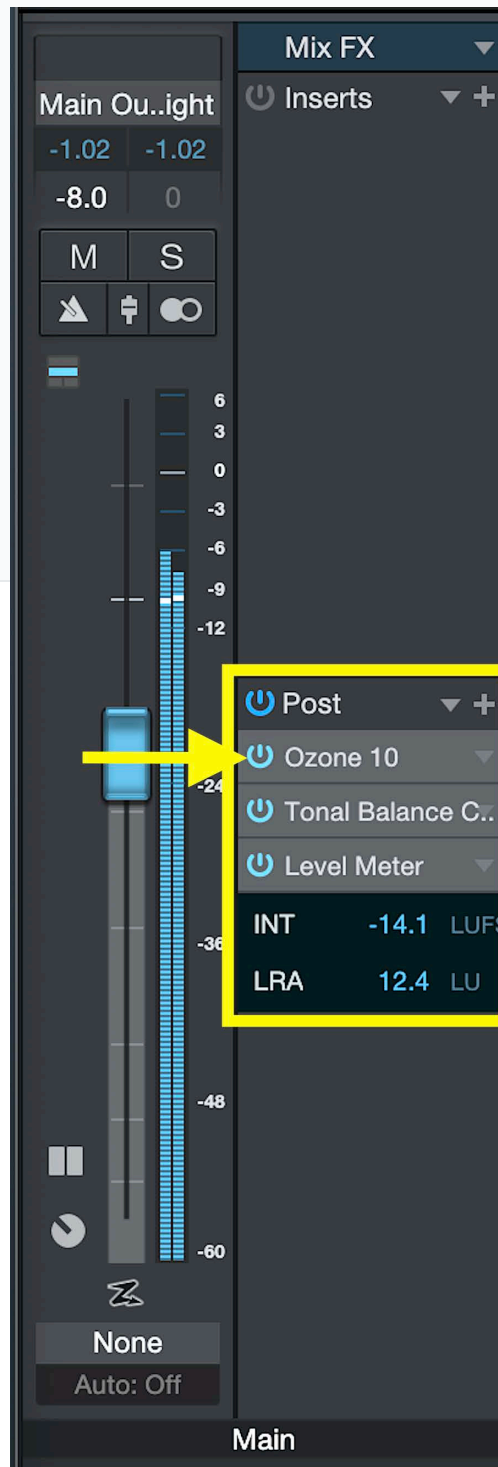


[Click on image to enlarge](#)

Adding Reverb to your mix puts listeners in a certain space, controlling where they are in relation to the music. Reverb is often described as adding warmth or 'wetness' to the sound. Less reverb on a track brings out the direct sound, bringing the instrument 'up front' and making it clear and present. More reverb on a track increases the depth of the sound, pushing the instrument 'to the rear' and making it distant and dreamy.

The Bach Adagio in D minor is somber, expressive, even a bit dark, with a melancholy melodic line over a slowly plodding basso continuo. This brings to mind being in a darkened, large open reverberant space when hearing this music. It is fitting, therefore, to set the Reverb parameters to reflect this listening environment. For this recording, the Lexicon MPX-i Reverb Preset "Dark Recital Hall" provides a reverb time of 1.7 seconds, 80% diffusion, and a dark coloration.

E3. Session 3: Mastering



In this session, the third of four, I'll talk about the mastering process in the creation of the musical recording of Bach's Adagio in D minor for keyboard solo. In a previous chapter, I gave an **overview of the mastering process**. The goal of mastering is to achieve the desired overall balance in spectral tone, dynamics, and stereo imaging in each song and between all the songs on an album. For an individual song, mastering offers a chance to fix up or tweak the balances that were set during the mixing process, and to bring up the volume to the desired sound level (volume normalization). For a set of songs, mastering provides consistent sound balancing across the whole set, so that the album has a unified sonic character.

In my home studio, I rely on my **PreSonus Studio One Professional** digital audio workstation (DAW) and the **iZotope Ozone 10 Mastering software** plug-ins to do the mastering. The iZotope Ozone 10 mothership plug-in is inserted post fader in the Main Output stereo channels of the DAW, as shown in the figure above.

The iZotope Ozone 10 mothership plug-in comprises a suite of signal processing plug-ins combined with artificial intelligence (AI) assistive audio technology for crafting a professional master. iZotope is one of the, if not *the*, leading signal processing software developers in the music production industry today. It is the “go to” source for many professional music studios, and I have come to appreciate how amazing these plug-ins are in a home music studio too.

The process begins with invoking the AI-powered Assistant to make a “first cut” putting together a signal processing chain of plug-ins based on your desired sound intent for tonal, spatial, and dynamic balances. This is shown below in the Assistant View Window of the Ozone 10 plug-in insert.



[Click on image to enlarge](#)

The user sets the intent controls of Tone Match, Width Match, and Dynamics Match, and the AI Assistant configures a chain of signal processing plug-ins that works toward achieving the desired tonal, width, and dynamic targets. The Tone Match intent control is connected to static and dynamic EQ plug-in modules. The Width Match intent control is connected to the Imager plug-in module. And the Dynamics Match intent control is connected to the Maximizer plug-in module.

Once the Assistant has created a “starting point” in your master, you can switch over to the Detailed View Window to see and make adjustments in the various plug-in modules.

The plug-in modules (EQ, Imager, Maximizer) that are used for mastering the piano tracks on this recording are shown below. The **overall tonal balance, stereo field width, and volume normalization** are set by the mastering process to achieve the musical sound I want for the Bach Adagio in D minor.



[Click on image to enlarge](#)

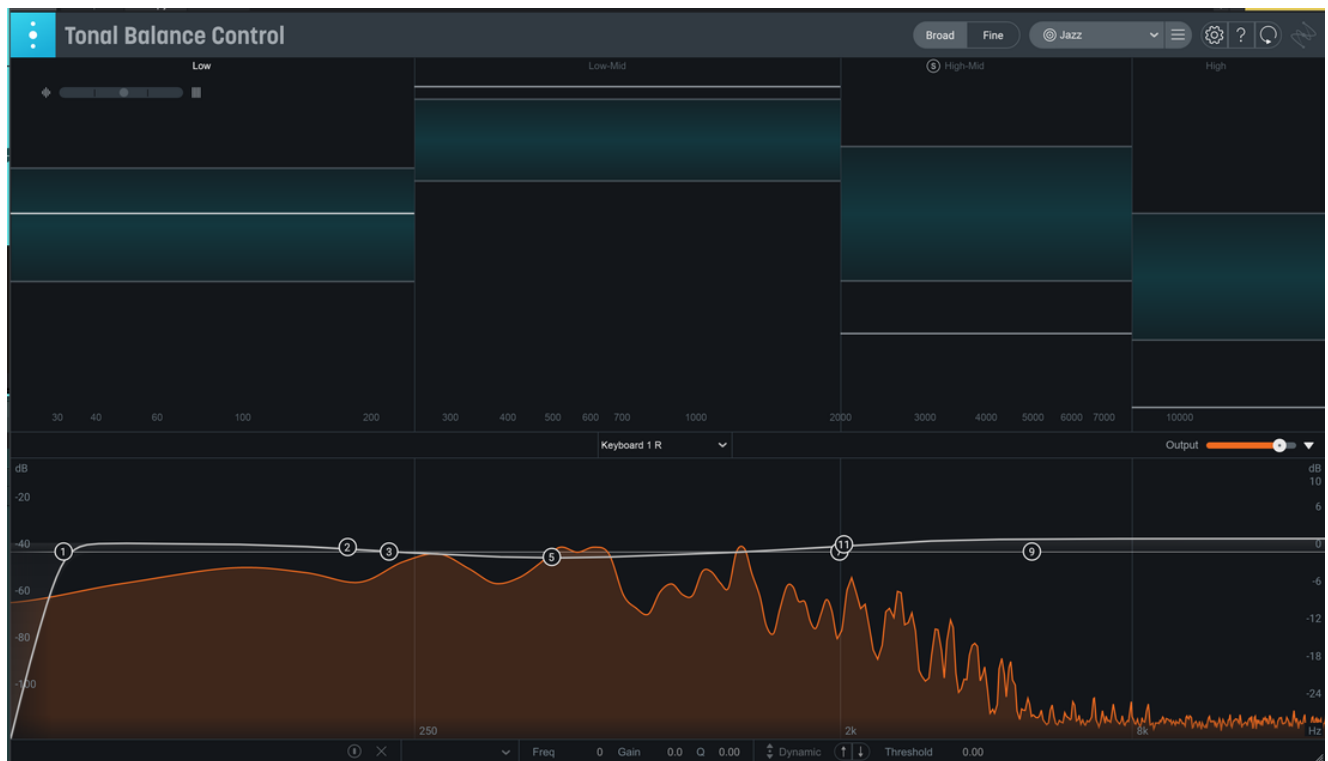


[Click on image to enlarge](#)



[Click on image to enlarge](#)

Finally, there are two metering plug-ins in the Main Output stereo channels – an iZotope Tonal Balance Control meter and a PreSonus Loudness Level (LUFS) meter. There is also a Vectorscope meter in the iZotope Imager plug-in (shown above) that provides a polar sample view of the stereo field image of the signal after all other signal processing is applied. These three meters serve as a second pair of “ears” to ensure that the mastering process successfully achieves the desired sound in your recording.



[Click on image to enlarge](#)



Time-Integrated Loudness Level (LUFS) Meter

In the next chapter, I'll talk about posting the Bach Adagio in D minor recording to a streaming platform, and making uncompressed, lossy, and lossless formatted sound files available to the public through Creative Commons licensing.

E4. Session 4: Streaming



In the last session in this series of four, I'll talk about uploading the mastered recording of Bach's Adagio in D minor to music hosting sites. From these host sites, the music can be streamed for public listening, or downloaded for public use according to its Creative Commons License.

When the mixing and mastering work in the digital audio workstation is complete, it's time to export (save) your stereo mix to a digital audio file for distribution online. For your audio file, you can choose from a number of different digital formats. These formats can be broadly categorized into three distinct groups: Uncompressed, Lossless, and Lossy. These three groups were discussed at length in the previous chapter on **Export Mixdown**. When I export a mixdown of a song, I do so creating several audio files with different formats. These audio files are destined for uploads to different hosting sites. Here's a table listing the audio files of my recording of J.S. Bach's Adagio in D minor for keyboard solo. The duration of the performed piece is 3 minutes, 40 seconds.

File	Format	Resolution	File Size	Host Site
Song.WAV	Uncompressed WAVE	24 bit 96 KHz	126.3 MB	Internet Archive archive.org/details/@pedalpointsound
Song.WAV	Uncompressed WAVE	24 bit 48 KHz	63.2 MB	Jamendo Music jamendo.com/artist/529848/gregory-tait
Song.FLAC	Lossless FLAC	24 bit 96 KHz Compression Level: 0 (least)	78.4 MB	Internet Archive archive.org/details/@pedalpointsound
Song.M4A	Lossless ALAC	24 bit 96 KHz	76.7 MB	Song download from Pedal Point Sound website pedalpointsound.com
Song.MP3	Lossy MP3	48 KHz sample rate 320 kbps bit rate	8.9 MB	Jamendo Music jamendo.com/artist/529848/gregory-tait

Here's a screen capture of my **Jamendo Music** artist page on a day when the Bach Adagio in D minor was at the top of the "Most Popular" list.

The screenshot displays the Jamendo Music artist page for Gregory Tait. The page layout includes a dark sidebar on the left with navigation links (Home, Explore, Radios, Spotlight) and a 'Jamendo Licensing' section. The main content area features a search bar, a 'LOG IN' button, and a large header image of a piano keyboard. Below the header, the artist's name 'Gregory Tait' is prominently displayed, along with his location 'USA' and genres '#folk, #classical, #christian'. A 'FOLLOW' button and a share icon are also present. To the right, statistics show '35,345 Listeners' and '7,051 Downloads'. The 'ARTIST' tab is selected, showing a list of tracks under the 'MOST POPULAR' filter. The first track, 'Bach Adagio in D minor' by Gregory Tait, is highlighted with a red box. Other tracks in the list include 'Fare Thee Well Sweet Ireland', 'Handel Harmonious Blacksmith Prelude', 'MacDowell Woodland Sketch 8', 'Abide with Me', 'Loch Lomond', 'MacDowell Woodland Sketch 1', 'Schumann Kinderszenen No. 12', and 'Ave Verum Corpus'. A 'PLAY ALL' button is located at the top right of the track list. To the right of the track list, there is a 'GREGORY TAIT BIOGRAPHY' section with a paragraph about his background and a 'Follow this artist' button.

Jamendo Music

Home
Explore
Radios
Spotlight

Are you an artist?
UPLOAD YOUR MUSIC

Search for independent music

LOG IN

Gregory Tait
USA #folk, #classical, #christian

FOLLOW

35,345 Listeners **7,051** Downloads

ARTIST **ALBUMS** **SINGLES** **SIMILAR ARTISTS** **COMMENTS**

MOST POPULAR **PLAY ALL**

Track Name	Artist	Actions
Bach Adagio in D minor	GREGORY TAIT	
Fare Thee Well Sweet Ireland	GREGORY TAIT	
Handel Harmonious Blacksmith Prelude	GREGORY TAIT	
MacDowell Woodland Sketch 8	GREGORY TAIT	
Abide with Me	GREGORY TAIT	
Loch Lomond	GREGORY TAIT	
MacDowell Woodland Sketch 1	GREGORY TAIT	
Schumann Kinderszenen No. 12	GREGORY TAIT	
Ave Verum Corpus	GREGORY TAIT	

GREGORY TAIT BIOGRAPHY

A classically trained pianist, Gregory Tait received his bachelor's degree in music and his doctoral degree in engineering. Drawing on a lifelong involvement in music and audio engineering, he has set up a home studio for performing and recording musical works. Current projects include recording short musical pieces from the classical, Christian, and folk repertoire. Further information is available at www.PedalPointSound.com

Follow this artist

Jamendo Licensing
Quality music for commercial use
START NOW!

Licensing Developers
Help center Contact us
Apps About us Jobs
Legal

English

Click image to enlarge

And here's a screen capture of my **Internet Archive** page, showing the various audio file formats available for download. The Creative Commons Attribution 4.0 International License (CC-BY) is prominently displayed.


INTERNET ARCHIVE

WEBBOOKSVIDEOAUDIOSOFTWAREIMAGES


ABOUTBLOGPROJECTSHelpDONATECONTACTJOBSVOLUNTEERPEOPLE





SIGN UP | LOG IN | UPLOAD


Search





Play from: Player Beta Webamp




 03:39 / 03:39 

 **Bach Adagio In D Minor**
by [Gregory Tait](#)

 Edit

 Manage



 History

Publication date

Usage

Topics

2023

Attribution 4.0 International  

Piano, Baroque Music, J.S. Bach, Keyboard Solo

In this recording, I perform Johann Sebastian Bach's Adagio in D minor for solo keyboard. This piece is very popular, and has been used in the soundtracks of several films, including Coda (2020) and Fifty Shades of Grey (2015) . This Baroque-era work is emotional, meditative, even a bit dark, with a beautiful upper melodic line undergoing many stylistic ornamentations over a slowly plodding, steady progression of chords in the low-mid register. The Adagio is a solo keyboard arrangement of the second movement of Bach's Concerto in D minor, BWV 974, which in turn is a transcription of the Concerto for Oboe and Strings in D minor composed by Alessandro Marcello. Bach's autograph of this keyboard arrangement has been lost, but the arrangement was copied around 1715 by Bach's second cousin, Johann Bernhard. It is this copy that we use today to perform the Adagio.

Addeddate

Identifier

Scanner

2023-03-13 22:22:24

bach-adagio-in-d-minor

Internet Archive HTML5 Uploader 1.7.0

Unfavorite

Share

Flag

22 Views

1 Favorite

DOWNLOAD OPTIONS

FLAC

1 file

ITEM TILE

1 file

TORRENT

1 file

VBR M3U

1 file

VBR MP3

1 file

WAVE

1 file


SHOW ALL

10 Files


6 Original

IN COLLECTIONS

Community Audio



Community Collections



Uploaded by

[PedalPointSound.com](#)

on March 13, 2023

Click image to enlarge

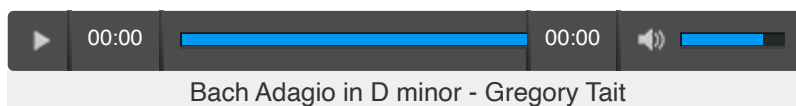
Finally, here's the uncompressed WAVE audio file (.wav) available for streaming and downloading from the Internet Archive website. At last, you can now listen to the "final product" that has been the subject of these last four sessions.

J.S. Bach

Adagio in D Minor

Performed by Gregory Tait

© 2023. This work is licensed under a [CC BY 4.0 license](#)



Appendix: Music Production Education

Music Technology Education



Music Production Studio

Choosing to take academic courses, perhaps even leading to a degree, in sound engineering and music production requires quite a bit of thought and planning. You need to ask the following questions:

Why should you attend a music production school or program ?

What knowledge and skills are you looking to obtain ?

What are your career options after earning your degree or certificate ?

Who are the top schools for music production ?

Music production programs are becoming very popular at colleges and universities across the nation. You can major in music production ! Programs have different names, such as Recording Arts, Music Technology, Audio Engineering, Sound Production and Recording, and Music Production. Some schools exclusively offer music production degrees, such as Berklee College of Music and Full Sail University. There are schools that focus on music technology, leading to a bachelor of science (B.S.) degree, while others are more music or music-business focused, leading to a bachelor of arts (B.A.) degree, a bachelor of fine arts (B.F.A.) degree, or a bachelor of music (B.M.) degree. There are masters-level graduate programs too.

Listed here are some of the big names in music production programs:

Berklee College of Music (B.M. Music Production and Engineering)

Full Sail University (B.S. Recording Arts, Audio Production;
certificate programs)

Musicians Institute (B.M. Composition (film score),
A.A.S. Studio Recording)

NYU Steinhardt (B.M. Music Technology)

USC Thornton School of Music (B.M. Music Production)

Drexel University Westphal College of Media Arts & Design
(B.S. Music Industry)

The knowledge and skills obtained from the technology portions of these programs include:

- Gaining expert experience using a Digital Audio Workstation and sound engineering electronic technology
- Learning how to use a sound mixer console, and training in fundamental principles of Mixing techniques
- Developing and implementing effective Mastering techniques

I received a B.A. degree in Music and Physics, and a Ph.D. degree in electrical engineering. My professional career was not involved with the music business. My experience in sound engineering and music production was gained through years of “doing sound” at church and through the process of creating and running a home music recording studio. I thought it would be instructive and fun to take a couple of online courses in music production to see how my development measured up against the knowledge and skills taught in these courses. I chose to take two certificate courses taught by Berklee School of Music faculty and offered through **Coursera** online.

The first certificate course I took was **The Technology of Music Production** from the Berklee School of Music. Here is an excerpt of the course description as printed in the school catalog :

“Learn about the music production process—including recording, editing, and mixing—and the tools available to you to create contemporary music on your computer.

With the recent introduction of high-quality-low-cost software and hardware, the tools of music production are now available to the masses. Albums are made in bedrooms as well as studios. On the surface this is liberating. Anyone can make an album for the low cost of a couple pieces of gear and a software package. But, if you dig deeper, you will find that it is not so easy. Producing music requires knowledge, dedication, and creativity.

Knowledge is where this course comes in. No matter what kind of music you are making, there is a large set of tools that you will need to use. Each lesson of this course will demonstrate a different set of music production tools, loosely following along the music production process of recording, editing, and mixing. We will start with some background on the nature of sound and how we perceive it. We will then examine the components necessary to record audio into a computer, so that you understand the devices that sound must travel through in a music production process. Once recorded, sound must be organized along a timeline, a process known as editing. It allows us to give the impression of perfect performances and create many of the sounds we hear in contemporary music. The contemporary editing tool is the Digital Audio Workstation (DAW), a piece of software that stores and organizes all the assets of a musical project. We will focus on the editing tools that are essential in contemporary music production and that all DAWs provide. After editing, sounds must be combined or mixed together, so we look to the mixing board—a very creative place if you know how to use it. We will explore the basic functionality of both hardware and software mixing boards, including volume, pan, mute, solo, buses, inserts, sends, and sub-mixes. The mixing process, however, includes more tools than the mixing board provides on its own. Sound must also be processed, modified from its recorded state to fit the context of the music. We will look at compression, equalization, and reverb, and examine the many audio effects that are offshoots of these devices and how they are used in a musical context.

In the end, the music production process relies on your creativity. Creativity is a product of the mind and will stay there, unexpressed, until the right tools are used in the right way to share it with the world. If you have an idea in your head, it will take numerous steps, each with an important tool, to reach your audience. You bring the dedication and creativity, and this course will bring you the knowledge to make that happen.”

The second certificate course I took was [The Art of Music Production](#) from the Berklee School of Music. Here is an excerpt of the course description as printed in the school catalog :

“Explore the art of record production and how to make recordings that other people will love listening to. This course will teach you how to make emotionally moving recordings on almost any recording equipment, including your phone or laptop. The emphasis is on mastering tangible artistic concepts; the gear you use is up to you. You will learn to develop the most important tool in the recording studio: your ears. You will learn to enhance every aspect of your own productions, both sonically and musically, by employing deeper listening skills.

Assignments will include posting your own recordings for peer review, and reviewing your classmates' work by employing specific tools and strategies. If you use a Digital Audio Workstation (DAW) to record and mix, that's great, but as long as you can record into your computer and post an MP3, you can complete the assignments. As you learn about the art of record production in this 4-week course, you will also learn about yourself and who you are as an artist and producer.”

There are two more certificate courses offered by Berklee School of Music that, together with the two discussed above, constitute a program in **Music Production Specialization**. These two additional courses are **Pro Tools Basics** and **Music Production Capstone**.

Given that the first two courses, the Technology of Music Production and the Art of Music Production, were very informative and fun, I would have liked to take these last two courses. But I am "married" to the PreSonus Studio One Pro digital audio workstation (DAW), rather than the Avid Pro Tools DAW, and I didn't want to take the time to climb the steep learning curve of the Pro Tools software. Too bad -- I'm sure I would have loved the capstone project. Anyway, I would highly recommend these online short courses from Berklee School of Music. And definitely worth noting - Berklee offers semester-length online courses in **Music Production** that come very highly recommended. They are well worth the time, effort, and dollars. If you are serious about a career in music production, these educational resources are a great place to start.